PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07H 21/00, G01N 33/53

A1

(11) International Publication Number: WO 92/13871

(43) International Publication Date: 20 August 1992 (20.08.92)

(21) International Application Number: PCT/US92/00863

(22) International Filing Date: 31 January 1992 (31.01.92)

(30) Priority data:
648,781 31 January 1991 (31.01.91) US

(71) Applicant: WASHINGTON UNIVERSITY [US/US]; One Brookings Drive, St. Louis, MO 63130 (US).

(72) Inventors: NORMARK, Staffan ; 30 Ridgemoor Drive, Clayton, MO 63105 (US). JONSSON, Ann-Beth ; Stipendiegrand 4C, S-901 01 Umea (SE).

(74) Agents: MONROY, Gladys, H. et al.; Morrison & Foerster, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, RU, SD, SE, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

Published
With international search report.

(54) Title: POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA

(57) Abstract

We have isolated and characterized a novel protein of pathogenic forms of *Neisseria*. We have also isolated and characterized genes which encode PilC, i.e., the *pilC* loci. Portions of the DNA sequences of the *pilC* genes are useful as probes to diagnose the presence of the relevant microorganisms containing type 4 pilin, for example, *Neisseria* in samples. These DNAs also make available polypeptide sequences of immunoreactive epitopes encoded within the loci, thus permitting the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or as components of vaccines. Antibodies, both monoclonal and purified polyclonal, directed against PilC epitopes are also useful for diagnostic test and as therapeutic agents for passive immunization. In addition, by utilizing probes derived from the DNA sequences, it is possible to isolate and sequence other portions of the *pilC* loci from species and strains of interest.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA

Technical Field

The invention relates to materials and methodologies for managing the spread of infections caused by microorganisms having type 4 pilin, for example, *Neisseria*. More specifically, it relates to polypeptides and antibodies useful in vaccines for the treatment of pathologic infections caused by these microorganisms. It also relates to polynucleotides useful for the recombinant production of these polypeptides. In addition, it relates to polypeptides, antibodies, and polynucleotides used for the detection of these strains.

Background Art

Type 4 pilins are expressed by several bacterial genuses, including *Neisseria, Moraxella, Bacteroides, and Pseudomonas*. Species within these genuses which have pathogenic members that express type 4 pilins are, for example, *N. gonorrhoeae, N. meningitidis, M. bovis, B. nodosus*, and *P. aeruginosa*. In addition, the Tcp pilin of *V. cholerae* is highly homologous to the type 4 pilins of other genuses.

The only known reservoir of the neisseriae is man. The genus includes two gram-negative species of pyogenic cocci that are pathogenic for man: the meningococcus (*Neisseria meningitidis*) and the gonococcus (*Neisseria gonorrheae*).

N. Meningitidis causes a variety of infections, most notably, meningitis and bacteremia. Meningococci can be divided into serologic groups on the basis of agglutination reactions with immune serum. The present classification includes groups A through Z. Clinically significant new groups encompass Y and W 135. The major groups are remarkably heterogeneous, but subclassification with additional serologic markers has been possible. Noncapsular antigens have provided the basis for dividing strains of groups into distinct types.

Meningococci cause either epidemic or sporadic disease, and historically, there has been a cyclic variation in the prevalence of meningococcal infection with peaks of increased frequency occurring every 8 to 12 years and lasting 4 to 6 years. The attack rate of meningococcal disease is highest for children between 6 months and 1 year. In the first half of this century, most epidemics of meningococcal disease in the United States were caused by group A organisms. In the past two decades, first group B then group C meningococci were responsible for outbreaks in both the military and civilian populations. Currently, group B is responsible for 50 to 55 percent of reported cases.

Gonorrhea, which is caused by *N. gonorrhea*, is an infection of columnar and transitional epithelium. This disease is the most common reportable communicable disease in the United States, and also has world-wide prevalence.

Although treatment of disease caused by gonococci and meningococci are often treated with antibiotics, these microorganisms often develop antibiotic resistance. Thus, prevention with vaccines is a preferable mode to contain the spread of infection. However, for a variety of reasons, including antigenic variation, the development of vaccines has been greatly hampered. For example, a vaccine which prevents gonorrhea is still lacking. In addition, although 56% of the causes of meningococcal disease are caused by serogroup B, an effective vaccine against this serogroup is also lacking.

N. gonorrheae and *N. meningitidis* are organisms completely adapted to the human host, having no other ecological niche. They have acquired a large arsenal of strategies to overcome the human host defense system.

The first step in infection with pathological forms of these *Neisseria* is adherence to target cells. It is thought that the pili of these microorganisms are a major virulence factor. For example, it is known that in the case of *N. gonorrheae*, piliated (P^+) variants attach much better to susceptible cells than non-piliated (P-) variants (Swanson, 1973; Pearce and Buchanan, 1978). Moreover, P+ variants, unlike P- variants, are able to establish an infection in human volunteers (Kellog et al, 1968).

Although the pilus protein elicits an immune response, so many antigenic variants exist and continue to develop that vaccines against the pilus protein are not highly effective.

Pilin is the major subunit of the pilus. Expression of pilin is controlled at the *pilE* locus.

Brief Description of the Invention

We have isolated and characterized a novel protein of pathogenic forms of *Neisseria*, PilC, that may be is associated with the pili of gonococci and meningococci. We have also isolated and characterized genes which encode PilC, i.e., the *pilC* loci.

Portions of the DNA sequences of the *pilC* genes are useful as probes to diagnose the presence of the relevant *Neisseria* in samples. These DNAs also make available polypeptide sequences of immunoreactive epitopes encoded within the loci, thus permitting the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or as components of vaccines for microorganisms with type 4 pilin and containing one or more epitopes that are immunologically identifiable with an epitope encoded in *pilC* of *Neisseria*. Antibodies, both monoclonal and purified polyclonal, directed against PilC epitopes are also useful for diagnostic tests and as therapeutic agents for passive immunization. In addition, by utilizing probes derived from the DNA sequences, it is possible to isolate and sequence portions of the *pilC* loci from species and strains of interest.

Accordingly, one embodiment of the invention is a recombinant polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Another embodiment of the invention is a recombinant expression system comprising a polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*, wherein the polynucleotide is operably linked to a control sequence compatible with a desired host.

Yet another embodiment of the invention is purified polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Another embodiment of the invention is a recombinant polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Still another embodiment of the invention is a vaccine composition for the treatment of *Neisseria infection*, comprised of a pharmaceutically acceptable excipient and of an effective amount of a recombinant polypeptide, wherein the polypeptide is comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Yet another embodiment of the invention is a composition comprised of purified polyclonal anti-PilC antibodies, wherein the PilC is of *Neisseria*.

An additional embodiment of the invention is a composition comprised of a monoclonal antibody directed against an immunoreactive epitope encoded in *pilC* of *Neisseria*.

Another embodiment of the invention is a method for producing antibodies to PilC of *Neisseria* comprising administering to an individual a composition comprised of an isolated immunogenic polypeptide containing a PilC epitope in an amount sufficient to produce an immune response.

Yet another embodiment of the invention is an oligomer capable of hybridizing to a sequence in *pilC* of *Neisseria*, wherein the oligomer is comprised of a *pilC* sequence complementary to at least about 6 contiguous nucleotides of *pilC*.

Still another embodiment of the invention is a process for detecting a *pilC* sequence in an analyte

strand, wherein the *pilC* sequence comprises a selected target region, the process comprising:

(a) providing a sample comprised of an analyte strand suspected of containing a selected target *pilC* sequence;

(b) providing an oligomer capable of hybridizing to the target *pilC* sequence, wherein the oligomer is comprised of a *pilC* targeting sequence complementary to at least about 6 contiguous nucleotides of *pilC*;

(c) incubating the sample of (a) with the oligomer of (b) under conditions which allow specific hybrid duplexes to form between the targeting sequence and the target sequence; and

(d) detecting hybrids formed between the target sequence, if any, and the oligomer.

Yet another embodiment of the invention is a recombinant polynucleotide comprising a DNA sequence of at least 8 contiguous nucleotides from *pilC*, wherein the *pilC* sequence is selected from the group of sequences shown in Figure 3, Figure 6, and Figure 7.

Another embodiment of the invention is a method of treating an individual for a *Neisseria* infection comprising administering to the individual antibodies produced according to claim 31, wherein the antibodies are administered in an amount effective to prevent the pathology of the infection.

Brief Description of the Drawings

Figure 1 is a genetic and physical map of *pilC* locus 1, showing the restriction enzyme sites.

Figure 2 shows the nucleotide sequence and the deduced amino acid sequence of the 5'-end of *pilC1*.

Figure 3 shows the nucleotide sequence of the sense strand of the *pilC1* gene.

Figure 4 shows the shows the nucleotide sequence of the sense strand of the *pilC1* gene and the amino acids encoded therein.

Figure 5 shows the nucleotide sequence of the sense strand of the *pilC1* gene, and the effect of frame shift on the putative gene products encoded therein.

Figure 6 shows the nucleotide sequences of PCR amplified fragments demonstrating a variation in the length of the G tract and sequence differences in the 5' region of the *pilC* genes.

Figure 7 shows the DNA sequence of the 3'-end of the *pilC2* fragment.

Figure 8 shows the *pilC2* fragment sequence, and the putative amino acids encoded therein.

Figure 9 shows a comparison of the analogous portions of *pilC2* (top) and *pilC1* (bottom) DNA sequences, and the putative amino acids encoded therein.

Modes for Carrying Out the Invention

The present invention provides polypeptides, antibodies, and polynucleotides which are useful for the detection and treatment of pathogenic microorganisms having type 4 pilin, for example, *Neisseria*, *Moraxella*, *Bacteroides*, and *Pseudomonas*.

We have discovered a polypeptide, PilC, which is present in *N. gonorrheae*. This polypeptide is a 110 kd protein that is closely associated with the pili of the microorganism. Most strains of *N. gonorrheae* carry two copies of the corresponding genes which encode the polypeptide(s); these genes have been denoted *pilC*. Expression from the *pilC* loci is regulated by frequent frameshift mutations within a run of G residues in the

region encoding the signal peptide. The two *pilC* genes of *N. gonorrheae* are not identical. Hence, alternate expression from either the *pilC1* or the *pilC2* loci gives rise to two different forms of PilC. Among nonpiliated (P^-) descendants from P^+ clones, clones were found that expressed pilin but not PilC. All P^+ revertants from such $PilC^-$ non-piliated clones have regained expression of PilC. Hence, phase variation of gonococcal pili can be caused by frameshift mutations in *pilC*. Transposon inactivation of the expressed *pilC2* copy resulted in a nonpiliated, pilin producing revertible phenotype. It appears, therefore, that PilC is required for assembly of pilin subunits into a polymerized pilus fiber in *N. gonorrheae*.

We have cloned and isolated gene, *pilC1*, from *N. gonorrheae*. In addition, by comparison of this gene sequence with a related sequence, we have cloned a fragment of the *pil2* gene. Moreover, using polynucleotide probes derived from isolated *pilC1* and PCR amplification, we have detected two possible variants of a *pilC* gene in *N. meningitidis*. The sequences of *pilC* reported herein appear to be novel, in that there are no reported counterparts in Genbank, and no significant homologies were found with any available sequences in that data base.

The useful materials and processes of the present invention are made possible by the provision of the sequences of the *pilC* genes from *N. gonorrheae* and from *N. meningitidis*. Information present in the sequences of the *pilC* genes allows for the design of polypeptides which may be useful as vaccines for treatment of pathogenic *Neisseria*, as diagnostic tools for the detection of these microorganisms, and as agents for the preparation of antibodies to PilC. In addition,

this information allows for the design of polynucleotides for the recombinant production of the polypeptides derived from PilC, and for the design of oligomers which are useful as probes and primers for the detection and amplification of target regions of *pilC*.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, MOLECULAR CLONING; A LABORATORY MANUAL, Second Edition (1989); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

As used herein, a polynucleotide "derived from" a designated sequence refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6 nucleotides, preferably at least about 8 nucleotides, more preferably at least about 10-12 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding to a region of the designated nucleotide sequence. "Corresponding" means homologous to or complementary to the designated sequence. Regions from which typical polynucleotide sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

The derived polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be consistent with an intended use.

Similarly, a polypeptide or amino acid sequence "derived from" a designated nucleic acid sequence refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence. This terminology also includes a polypeptide expressed from a designated nucleic acid sequence.

A recombinant or derived polypeptide is not necessarily translated from a designated nucleic acid

sequence. It may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from a microorganism. A recombinant or derived polypeptide may include one or more analogs of amino acids or unnatural amino acids in its sequence. Methods of inserting analogs of amino acids into a sequence are known in the art. It also may include one or more labels, which are known to those of skill in the art.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, (3) does not occur in nature, or (4) is not in the form of a library.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example proteins (including for e.g., nucleases, toxins, antibodies, signal peptides,

poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

The term "purified polynucleotide" refers to a polynucleotide which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90% of polypeptides with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides from bacteria are known in the art, and include for example, disruption of the bacteria with a chaotropic agent, differential extraction and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

The term "purified polypeptide" refers to a polypeptide or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90%, of cellular components with which the polypeptide is naturally associated. Techniques for purifying polypeptides are known in the art, and examples of these techniques are discussed infra.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in

morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc. that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers. The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this

region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

"Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are also present in the designated polypeptide(s). Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art, and are also illustrated infra.

As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance.

A polypeptide is "immunoreactive" when it is "immunologically reactive" with an antibody, i.e., when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody

binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an epitope against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art. An "immunoreactive" polypeptide may also be "immunogenic". As used herein, the term "immunogenic polypeptide" is a polypeptide that elicits a cellular and/or humoral immune response, whether alone or linked to a carrier in the presence or absence of an adjuvant.

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one antibody combining site. An "antibody combining site" or "binding domain" is formed from the folding of variable domains of an antibody molecule(s) to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows an immunological reaction with the antigen. An antibody combining site may be formed from a heavy and/or a light chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, altered antibodies, univalent antibodies, the Fab proteins, and single domain antibodies.

As used herein, a "single domain antibody" (dAb) is an antibody which is comprised of an VH domain, which reacts immunologically with a designated antigen. A dAB does not contain a VL domain, but may contain other antigen binding domains known to exist in antibodies, for example, the kappa and lambda domains. Methods for preparing dABs are known in the art. See, for example, Ward et al. (1989).

Antibodies may also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation are known in the art (see, e.g., U.S. Patent No. 4,816,467, which is incorporated herein by reference), and include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of all the chains of a particular antibody are homologous with the chains found in one antibody produced by the lymphocyte which produces that antibody *in situ*, or *in vitro* (for example, in hybridomas). Vertebrate antibodies typically include native antibodies, for example, purified polyclonal antibodies and monoclonal antibodies. Examples of the methods for the preparation of these antibodies are described infra.

"Hybrid antibodies" are antibodies wherein one pair of heavy and light chains is homologous to those in a first antibody, while the other pair of heavy and light chains is homologous to those in a different second antibody. Typically, each of these two pairs will bind different epitopes, particularly on different antigens. This results in the property of "divalence", i.e., the ability to bind two antigens simultaneously. Such hybrids may also be formed using chimeric chains, as set forth below.

"Chimeric antibodies", are antibodies in which the heavy and/or light chains are fusion proteins. Typically the constant domain of the chains is from one particular species and/or class, and the variable domains

are from a different species and/or class. Also included is any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be differing classes, or different species of origin, and whether or not the fusion point is at the variable/constant boundary. Thus, it is possible to produce antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

Another example is "altered antibodies", which refers to antibodies in which the naturally occurring amino acid sequence in a vertebrate antibody has been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region may be made to alter antigen binding characteristics. The antibody may also be engineered to aid the specific delivery of a molecule or substance to a specific cell or tissue site. The desired alterations may be made by known techniques in molecular biology, e.g., recombinant techniques, site directed mutagenesis, etc.

Yet another example are "univalent antibodies", which are aggregates comprised of a heavy chain/light chain dimer bound to the Fc (i.e., constant) region of a second heavy chain. This type of antibody escapes antigenic modulation. See, e.g., Glennie et al. (1982).

Included also within the definition of antibodies are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion . "Fab" includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as $F(ab)_2$), which are capable of selectively reacting with a designated antigen or antigen family. "Fab" antibodies may be divided into subsets analogous to those described above, i.e, "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing "Fab" fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

The term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as

other modifications known in the art, both naturally occurring and non-naturally occurring.

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

"Treatment" as used herein refers to prophylaxis and/or therapy.

By "immunogenic" is meant an agent used to stimulate the immune system of a living organism, so that one or more functions of the immune system are increased and directed towards the immunogenic agent. Immunogenic agents include vaccines. Immunogenic agents can be used in the production of antibodies, both isolated polyclonal antibodies and monoclonal antibodies, using techniques known in the art.

By vaccine is meant an agent used to stimulate the immune system of a living organism so that protection against or amelioration of future harm is provided. Immunization refers to the process of inducing an increased level of antibodies and/or cellular immune response in which T-lymphocytes respond by killing the pathogen and/or activate other cells involved in the immune response pathway. The antibodies produced as a result of immunization may belong to any of the immunological classes, such as immunoglobulins A, D, E, G or M.

An "individual", as used herein, refers to vertebrates, particularly members of the mammalian

species, and includes but is not limited to domestic animals, sports animals, and primates, including humans.

As used herein, the "sense strand" of a nucleic acid contains the sequence that has sequence homology to that of mRNA. The "anti-sense strand" contains a sequence which is complementary to that of the "sense strand".

As used herein, the term "probe" refers to a polynucleotide which forms a hybrid structure with a sequence in a target region, due to complementarity of at least one sequence in the probe with a sequence in the target region. The polynucleotide regions of probes may be composed of DNA, and/or RNA, and/or synthetic nucleotide analogs.

As used herein, the term "target region" refers to a region of the nucleic acid which is to be amplified and/or detected. The term "target sequence refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The term "primer" as used herein refers to an oligomer which is capable of acting as a point of initiation of synthesis of a polynucleotide strand when placed under appropriate conditions. The primer will be completely or substantially complementary to a region of the polynucleotide strand to be copied. Thus, under conditions conducive to hybridization, the primer will anneal to the complementary region of the analyte strand. Upon addition of suitable reactants, (e.g., a polymerase, nucleotide triphosphates, and the like), the primer is extended by the polymerizing agent to form a copy of the analyte strand. The primer may be single-stranded, or alternatively may be partially or fully double-stranded.

The terms "analyte polynucleotide" and "analyte strand" refer to a single- or double-stranded nucleic

acid molecule which is suspected of containing a target sequence, and which may be present in a biological sample.

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of *in vitro* cell culture constituents (including but not limited to conditioned medium resulting from the growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

As used herein, the term "oligomer" refers to primers and to probes. The term oligomer does not connote the size of the molecule. However, typically oligomers are no greater than 1000 nucleotides, more typically are no greater than 500 nucleotides, even more typically are no greater than 250 nucleotides; they may be no greater than 100 nucleotides, and may be no greater than 75 nucleotides, and also may be no greater than 50 nucleotides in length.

The term "coupled" as used herein refers to attachment by covalent bonds or by strong non-covalent interactions (e.g., hydrophobic interactions, hydrogen bonds, etc.). Covalent bonds may be, for example, ester, ether, phosphoester, amide, peptide, imide, carbon-sulfur bonds, carbon-phosphorus bonds, and the like.

The term "support" refers to any solid or semi-solid surface to which a desired polypeptide or polynucleotide may be anchored. Suitable supports include glass, plastic, metal, polymer gels, and the

like, and may take the form of beads, wells, dipsticks, membranes, and the like.

The term "label" as used herein refers to any atom or moiety which can be used to provide a detectable (preferably quantifiable) signal, and which can be attached to a polynucleotide or polypeptide.

The term "type 4 pilin" as used herein refer to pilins that contain a conserved amino terminal hydrophobic domain beginning with an amino-terminal phenylalanine that is methylated upon processing and secretion of the pilin. Another characteristic feature of type 4 pilins is that in the propilin form they contain similar six- or seven-amino acid long leader peptides, which are much shorter than typical signal sequences. Type 4 pilins are expressed by several bacterial genuses, including *Neisseria, Moraxella, Bacteroides, and Pseudomonas*. Species within these genuses which express type 4 pilins are, for example, *N. gonorrhoeae, N. meningitidis, M. bovis, B. nodosus*, and *P. aeruginosa*. As used herein, the term "type 4 pilin" also includes the Tcp pilin of *Vibrio*, (for example, V. cholerae), that is highly homologous to the type 4 pilins of other genuses. Tcp pilin contains the characteristic amino-terminal hydrophobic domain as well as having a modified N-terminal amino acid that in this case may be a modified methionine because the Tcp pilin gene encodes a methionine residue at the position where all the others encode a phenylalanine. Precursor TcpA contains a much longer leader sequence than typical type 4 propilins but retains homology in the region surrounding the processing site.

The term "*pilC*" as used herein refers to a gene encoding a polypeptide involved in the assembly of type 4 pilin, which may also be required for attachment of the

pilin, and that is comprised of an epitope that is immunologically identifiable with an epitope in PilC of *N. gonorrhae* or *N. meningitidis*. Included within this term is any homologous region from *Vibrio*, *tcpC*.

As used herein the term "PilC" refers to a polypeptide encoded within *pilC*, and includes TcpC of *Vibrio*.

The description of the method to retrieve the DNA sequences is mostly of historical interest. The resultant sequences (and their complements) are provided herein, and the sequences, or any portion thereof, could be prepared using synthetic methods, or by a combination of synthetic methods with retrieval of partial sequences using methods similar to those described herein.

The description infra, of "walking" the genome by isolating overlapping DNA sequences from the *N. gonorrheae* lambda gt-11 library and from an EMBL3 library provides one method by which DNAs corresponding to the *pilC* genomes from, inter alia, *N. gonorrheae* and *N. meningitidis*, respectively, may be isolated. However, given the information provided herein, other methods for isolating *pilC* DNAs from these species, as well as from species of other genuses which have type 4 pilin are obvious to one of skill in the art.

Characterization of the genes of the *pilC* loci has provided information on the polypeptides encoded therein, and on the control of their expression. Even though Type 4 pili have been extensively studied in several laboratories, little is known about their assembly. The presence of a specific assembly machinery for this class of pili is evident from the fact that the pilin gene of *B. nodosus* and *M. bovis* can be properly processed and assembled into a pilus in *P. aeruginosa* but not in *E. coli* (Ellerman et al., 1986; Mattick et al.,

1987; Beard et al., 1990). Furthermore, the recent genetic characterization of TCP pili of *Vibrio cholerae* has revealed that a number of closely linked genes are required for pilin processing and assembly into a structure (Taylor et al., 1988). The TCP pilin does not carry an N-methylphenylalanine but its primary sequence is highly homologous to the Type 4 class of pilins.

The *N. gonorrhoeae* pilus facilitates adherence of the bacterium to a number of eukaryotic cell types (Watt et al., 1980) and is thought to play a role in bacterial interaction with neutrophils (Fischer and Rest, 1988). The pilin is encoded from one or two *pilE* loci (Meyer et al., 1984; Swanson et al., 1986) which most likely each form a monocistronic operon. Hence, there have been no suggestions that genes closely linked to *pilE* are involved in pilus assembly. A dispersed location of genes involved in gonococcal pilus assembly as well as the rapid occurrence of nonpiliated variants generated via recombination with pilin sequences from silent loci, *pilS*, have made it extremely difficult to identify putative assembly genes for gonococcal pili.

The PilC protein described herein is a protein encoded within a *pilC* or equivalent (for example, *tcpC*) locus or gene. In *N. gonorrhoeae* MS11 and most other gonococcal strains the PilC protein is expressed in small amounts. It is the only protein that is enriched in highly purified preparations of MS11 pili. PilC was not released from a nonpilated MS11 (P^-n) variant using the same procedure suggesting that this protein interacts with the polymerized pilus fiber.

DNA sequence analysis of the cloned *pilC1* gene revealed one long open reading frame that was out of frame with its putative AUG initiation codon and 5′ end encoding the signal peptide. Minute amounts of PilC were

expressed in *E. coli* from pABJ04. Gel purified PilC from MS11 contained a lysine residue in position four, whereas *pilC1* had a glutamine codon at this position. A lysine codon was, however, found at position four in a number of PCR amplified 5' *pilC* fragments suggesting that these fragments represent the 5' end of *pilC2*, which then must be ON in MS11. The finding that a miniTnCm insertion in *pilC2* abolished PilC expression, whereas insertional inactivation of *pilC1* did not abolish PilC expression further argues that *pilC1* is translationally out of frame and *pilC2* translationally in frame in the MS11 variant we are studying.

PCR amplified fragments of *pilC1* and *pilC2* in MS11 differed in the number of G residues found in the G tract. Only 11 or 12 Gs were found in *pilC1* clones (which would both generate an OFF phenotype) while 12 or 13 Gs were found among *pilC2* specific clones. Since *pilC2* is the expressed gene in the MS11 variant under study, we believe that this variant carries 13 Gs in *pilC2* and 12 Gs in *pilC1*. The frequency of frameshift mutations in each locus is not known. However, the lack of 13 Gs among *pilC1* specific fragments and the lack of 14 Gs among *pilC2* specific fragments suggests that a deletion of one G residue occurs at a higher frequency than the insertion of one G residue. We had expected to find amplified fragments from *N. gonorrhoeae* containing 10 G residues in the G tract, but found none in the 48 clones sequenced. If only one G is added or deleted in each mutational event, the frequency of G tracts with 10 residues should be low if G tracts normally are 12 or 13 bp long.

Frameshifting in *pilC1* also occurred in *E. coli*. In this case, however, two variants with 10 residues were found out of 12 clones sequenced. It may

therefore be that there is a selection against in frame variants with 10 Gs in *N. gonorrhoeae*. A change from five glycines to four in the signal peptide may for example have an effect on the physical properties of the precursor form of PilC such that the signal peptide is not cleaved off. *E. coli* strain AA10 is *rec*A. Therefore, frameshift mutations in the G tract of *pilC* occurs independent of the RecA protein.

Translational frameshifting has been shown to regulate phase and antigenic variation of the gonococcal opacity protein PII that is encoded by a number of opa loci showing sequence variations. In this system a number of pentameric CTCTT repeats are present in the region encoding the signal peptide (Stern et al. 1986). Variation in the number of repeats is independent of *rec*A in *N. gonorrhoeae* as well as in *E. coli* (Murphy et al., 1989). Variation in the expression of lipopolysaccharide epitopes in *Haemophilus influenzae* was recently explained by translational frameshifting created by alterations in the number of CAAT repeats occurring in the 5' end of *lic*A (Weiser et al., 1989). In *Bordetella pertussis* frameshift mutations in the regulatory *vir* locus occur in a run of C residues positioned internally in the gene (Stibitz et al., 1989). The C tract was in this case varying from 6 (in frame) to 7 residues (out of frame). It is not known if this frameshift mutation is programed or not. The pilin gene of *Bordetella pertussis* was recently shown to be preceded by a stretch of Cs. Frequent mutations affecting the length of this C tract influenced the transcriptional activity of the pilin gene (Willems et al., 1990).

Variation in the number of the CTCTT repeats in opa genes was recently suggested to be due to recombination-independent slipped strand mispairing

(Murphy et al., 1989). Mispairing is thought to occur between strands subjected to local denaturation and should preferentially occur during replication. A number of unusual DNA structures (cruciform, Z form, H form) have been shown to form *in vitro* within a variety of specific DNA sequences. Under normal conditions the B form is the most favorable thermodynamically (Frank-Kamenetskii and Vologodskii, 1984). Transition to alternative conformations requires specific external conditions, supercoiling being the most physiologic. Single stranded $(dG)_n$ and $(dC)_n$ strands renature more slowly than complementary strands with arbitrary sequences, and methylation experiments suggest that a poly dG chain may form a hairpin-like structure stabilized by G-G bp (Panyutin et al., 1990).

Four variant sequences differing outside the G tract were obtained by PCR amplification of the 5′ end of *pilC* from four *N. gonorrhoeae* strains. The region 5′ of the G tract was invariant, as was the 3′ end of the amplified region. All variation was confined to a region located 3′ of the G tract. At least some of these sequence variations can be explained by mismatch pairing events. Thus, the addition of four nucleotides distal to the G tract in variant sequence 4 is possible to explain by a two step mispairing event occurring within variant sequence 2. Slip strand mispairing between the two CA residues in -GGCGCAGGCGCA- would yield -GGCGCAGGCGCAGGCGCA-. A second mispairing event occurring between the two C-residues at positions 3 and 5 gives rise to the sequence -GGCAGGCGCAGGCGCA- present in variant 4. It may therefore be that a sequence close to a poly(G) tract is prone to slipped strand mispairing.

Gonococcal pilus phase variation is associated with an altered nucleotide sequence of *pilE* via

recombinations with silent *pilS* sequences (Haas and Meyer, 1986, Swanson et al., 1986). An irreversible switch OFF in pilus expression results from deletions of the 5′ coding and control regions of the *pilE* locus (Swanson et al., 1985). Reversible gonococcal pilus phase variation is associated with nucleotide changes in *pilE* resulting in an altered pilin product. It has been suggested that the pilins of these variants are assembly defective (Bergstrom et al., 1986; Swanson et al., 1986; Hill et al., 1990). Here we present evidence that switch OFF and ON of PilC expression causes pilus expression to phase vary. Five out of five P^-, pilin producing descendants from $MS11_{mk}$ (P^+, $PilC^+$) that expressed pilin did not express PilC. All tested P^+ revertants from the five P^-, $PilC^-$ variants had regained expression of PilC. The pilin of one nonpiliated PilC OFF-switcher (variant 8) differed by eight amino acids from that of the parent. The fact that one piliated $PilC^+$ backswitcher (8:1) expressed a pilin identical in sequence to the nonpiliated variant (8) strongly suggests that the regained expression of pili is due to an ON-switch in PilC expression. The above results also imply that the nonpiliated phenotype of variant 8 is not due to the alterations in the pilin relative to the parental strain but to an OFF-switch of PilC. The finding that mTnCm insertions resulted in P^+ colonies when inserted into *pilC1* and P^- colonies when inserted into the actively expressing *pilC2* locus offers further evidence that PilC is essential for the biogenesis of gonococcal pili. P^-, *pilC2*::mTnCm-12 insertion mutants reverted to P^+ colony morphology at a low frequency. These revertants most likely represent frameshifting mutants in *pilC1* resulting in expression of PilC from this locus. A double mutant in *pilC1* and *pilC2* was stably nonpiliated, expressed

pilin, but expressed amounts of pilin that did not express any pili when examined by transmission electron microscopy. It is therefore believed that out of frame mutations of both *pilC1* and *pilC2* will abolish pili formation.

At this stage we cannot exclude the possibility that some PilC$^-$ variants from MSII$_{mk}$(P$^+$,PilC$^-$) are generated by transformation of *pilC1* sequences and homologous recombination with *pilC2* thus generating variants with two *pilC1* 5' ends at both *pilC* loci. PilC$^+$ revertants from PilC$^-$ clones must, however, all be due to frameshift mutations in either *pilC1* or *pilC2*.

We propose that PilC forms an outer membrane pore or assembly center enabling the pilin subunits to be assembled and translocated across the outer membrane analogous to the proposed function of the high molecular weight proteins required for the assembly of enterobacterial pili (the latter of which is discussed in Norgren et al., 1987). Alternatively, PilC may act as an initiator for polymerization. In the latter case PilC would be expected to be located at the tip of the polymerized pilus.

It is possible that the alternate expression of PilC from two structurally different *pilC* loci is yet another example of antigenic variation in *Neisseria gonorrhoeae*. It is, however, possible that this variation could have functional implications as well. Each class of *E. coli* pili utilizes a different outer membrane pore/assembly protein. Hence, pilin subunits and/or periplasmic chaperone complexes may specifically interact with an exposed region of the protein allowing polymerization of pilus subunit proteins. The repertoire of antigenic variants of gonococcal pilins is vast (Hagblom et al., 1985). It may be that only certain

pilin variants are assembled via PilC1 and PilC2 respectively. Alternatively, if PilC acts as an initiator it could also possess other properties such as being involved in Pilus mediated attachment.

In one embodiment of the invention, immunogenically active polypeptides encoded within *pilC* are prepared. The availability of *pilC* DNA sequences, either those isolated by utilizing the DNA sequences described in the Examples, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding immunologically reactive regions of the polypeptide encoded in either strand. Immunological reactivity may be determined by immunoassay using antibodies raised to PilC. Fragments encoding the desired polypeptides are derived from the DNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986. Any desired portion of the *pilC* DNA containing an open reading frame, in either sense strand, can be obtained as a recombinant polypeptide, such as a mature or fusion protein; alternatively, a polypeptide encoded in the DNA can be provided by chemical synthesis.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host

systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell lines is given infra. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such polypeptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy.

The PilC antigens may also be isolated from meningococci and from gonococci. The bacteria may be grown by conditions known in the art, some of which are described infra. In addition, a method for isolating PilC from gonococci is described infra.

In another embodiment of the invention, the immunoreactive polypeptides may be conjugated with carrier. An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of PilC antigen. Accordingly, using the DNAs of *pilC* as a basis, DNAs encoding short segments of PilC polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the synthesized polypeptide is correctly configured so as to provide the correct

epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridyl-thio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun. Rev. (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized

macromolecules such as proteins; polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles or attenuated bacteria of other strains, for example, those of *Salmonella*. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

In addition to full-length PilC proteins, polypeptides comprising truncated PilC amino acid sequences encoding at least one immunologically reactive epitope are useful immunological reagents. For example, polypeptides comprising such truncated sequences can be used as reagents in an immunoassay. These polypeptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While these truncated sequences can be produced by various known treatments of native bacterial protein, it is generally preferred to make synthetic or recombinant polypeptides comprising a PilC sequence. Polypeptides comprising these truncated PilC sequences can be made up entirely of PilC sequences (one or more epitopes, either contiguous or noncontiguous), or PilC sequences and heterologous sequences in a fusion protein. Useful heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the PilC epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, e.g., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783, the disclosures of which are incorporated herein by reference.

The size of polypeptides comprising the truncated PilC sequences can vary widely, the minimum size being a sequence of sufficient size to provide an immunologically reactive PilC epitope, while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired PilC epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated PilC amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the PilC sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select PilC sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

Truncated PilC amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire PilC protein sequence can be screened by preparing a series of short peptides that together span the entire protein sequence. By starting with, for example, 100mer polypeptides, it would be routine to test each polypeptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified 100mer to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare oligopeptides comprising the identified regions for screening.

In another embodiment of the invention, the immunogenicity of the epitopes of PilC may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins

such as, for example, that associated with hepatitis B surface antigen. See, e.g., U.S. Pat. No. 4,722,840. Constructs wherein the PilC epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the PilC epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include PilC sequences are immunogenic with respect to the microorganism encoding the PilC epitope (for example, *Neisseria, Vibrio, Moraxella, Bacteroides, or Pseudomonas*) and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in *S. cerevisiae* (Valenzuela et al. (1982)), as well as in, for example, mammalian cells (Valenzuela, P., et al. (1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding a PilC epitope. In this replacement, regions which are not required to mediate the aggregation of the

units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the PilC epitope.

In another embodiment of the invention, the immunoreactive polypeptides encoded in *pilC* are prepared into vaccines. Vaccines may be prepared from one or more immunogenic polypeptides derived from *pilC*. If recombinant, these polypeptides may be expressed in various host cells (e.g., bacteria, yeast, insect, or mammalian cells), or alternatively may be isolated from the bacterial preparations. In addition to the above, it is also possible to prepare live vaccines of attenuated microorganisms which express one or more recombinant polypeptides derived from the *pilC* gene. Suitable attenuated microorganisms are known in the art and include, for example, viruses (e.g., vaccinia virus (see Brown et al. (1986)), as well as bacteria.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples

of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing a PilC immunoreactive sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic antigen(s) derived from *pilC* may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

Another embodiment of the invention are antibodies which react immunologically with PilC epitopes. The immunogenic polypeptides prepared as described above are used to produce antibodies, including polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing a PilC epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to a PilC epitope (i.e., an epitope encoded within *pilC*) contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Monoclonal antibodies directed against PilC epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against PilC epitopes can be screened

for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against PilC epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Methods for introducing antibodies into an individual to accomplish passive immunotherapy are known in the art. In addition, monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985). Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for treatment, vaccination and/or diagnosis of the relevant microorganism encoding the antigen of interest, (for example, *Neisseria*, *Pseudomonas*, *Moraxella*, *Bacteroides*, or *Vibrio*) as well as for an elucidation of the immunogenic regions of PilC.

Another embodiment of the invention concerns immunoassays and diagnostic kits. The polypeptides which contain epitopes encoded in *pilC* which are immunoreactive with anti-PilC antibodies in biological samples are useful in immunoassays to detect presence of anti-PilC antibodies, or the presence of the relevant microorganism or its antigens in biological samples. Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. The immunoassay will utilize a polypeptide comprised of at least one epitope derived from PilC or encoded in *pilC*. In one embodiment,

the immunoassay uses a combination of epitopes including the one derived from PilC or encoded in *pilC*. These epitopes may be derived from the same or from different polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides. An immunoassay may use, for example, a monoclonal antibody directed towards an epitope(s), a combination of monoclonal antibodies directed towards epitopes of one antigen, monoclonal antibodies directed towards epitopes of different antigens, polyclonal antibodies directed towards the same antigen, or polyclonal antibodies directed towards different antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Typically, an immunoassay for an anti-PilC antibody(s) will involve selecting and preparing the test sample suspected of containing the antibodies, such as a biological sample, then incubating it with an immunoreactive (also called antigenic) polypeptide(s) containing at least one epitope encoded in *pilC*. The incubation is under conditions that allow antigen-antibody complexes to form. Suitable incubation conditions are well known in the art. Subsequent to the incubation, complexes which are formed which contain the immunoreactive polypeptide are detected. The immunoassay

may be, without limitations, in a heterogenous or in a homogeneous format, and of a standard or competitive type.

In a heterogeneous format, the polypeptide is typically bound to a solid support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e.g., in membrane or microtiter well form), polyvinyl chloride (e.g., in sheets or microtiter wells), polystyrene latex (e.g., in beads or microtiter plates, polyvinylidine fluoride (known as Immulon), diazotized paper, nylon membranes, activated beads, and Protein A beads. For example, Dynatech Immulon[1] or Immulon[2] microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic polypeptide is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

In a homogeneous format, the test sample is incubated with antigen in solution. For example, it may be under conditions that will precipitate any antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of anti-PilC antibodies forming the antibody-antigen complex is directly monitored. This may be accomplished by determining whether labeled anti-xenogeneic (e.g., anti-human) antibodies which recognize an epitope on anti-PilC antibodies will bind due to complex formation. In a competitive format, the amount of anti-PilC antibodies in the sample is deduced by monitoring the

competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-PilC antibody (or, in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled anti-PilC antibodies in the complex may be detected using a conjugate of antixenogeneic Ig complexed with a label, (e.g., an enzyme label).

In immunoassays where PilC polypeptides are the analyte, the test sample, which may be a biological sample, is incubated with anti-PilC antibodies under conditions that allow the formation of antigen-antibody complexes. Various formats can be employed. For example, a "sandwich assay" may be employed, where antibody bound to a solid support is incubated with the test sample; washed; incubated with a second, labeled antibody to the analyte, and the support is washed again. Analyte is detected by determining if the second antibody is bound to the support. In a competitive format, which can be either heterogeneous or homogeneous, a test sample is usually incubated with antibody and a labeled, competing antigen is also incubated, either sequentially or simultaneously. These and other formats are well known in the art.

The antigenic regions of the polypeptides encoded in *pilC* can be mapped and identified by screening the antigenicity of expression products of *pilC* DNAs which encode portions of the PilC. The expression products may be from a variety of expression systems, including, for example bacterial systems, yeast systems, insect systems, and eukaryotic cell systems. In addition, studies giving rise to an antigenicity index

and hydrophobicity/hydrophilicity profile give rise to information concerning the probability of a region's antigenicity.

Efficient detection systems for infection with pathogenic microorganisms, (for example, *Neisseria, Pseudomonas, Bacteroides, Moraxella, or Vibrio*) may include the use of panels of epitopes. The epitopes in the panel may be constructed into one or multiple polypeptides. At least one of the epitopes will be encoded in *pilC* or derived from PilC. The assays for the varying epitopes may be sequential or simultaneous.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing PilC epitopes (i.e., epitopes encoded within *pilC*) or antibodies directed against PilC epitopes in suitable containers. The kit may also contain other reagents, for example, buffer and standard, as well as other materials required for the conduct of the assay, as well as a suitable set of instructions for conducting the assay using the kit materials.

Another embodiment of the invention are oligomers. Using the disclosed portions of the *pilC* DNAs as a basis, oligomers of approximately 8 nucleotides or more can be prepared, either by excision from recombinant polynucleotides or by synthetic methods which are known in the art. These oligomers can serve as probes for the detection (including isolation and/or labeling) of polynucleotides which contain *pilC* sequences, and/or as primers for the transcription and/or replication of targeted *pilC* sequences. The oligomers contain a targeting polynucleotide sequence, which is comprised of nucleotides which are complementary to a target *pilC*

nucleotide sequence; the sequence is of sufficient length and complementarity with the *pilC* sequence to form a duplex which has sufficient stability for the purpose intended. For example, if the purpose is the isolation, via immobilization, of an analyte containing a target *pilC* sequence, the oligomers would contain a polynucleotide region which is of sufficient length and complementarity to the targeted *pilC* sequence to afford sufficient duplex stability to immobilize the analyte on a solid surface, via its binding to the oligomers, under the isolation conditions. For example, also, if the oligomers are to serve as primers for the transcription and/or replication of target *pilC* sequences in an analyte polynucleotide, the oligomers would contain a polynucleotide region of sufficient length and complementarity to a region flanking the targeted *pilC* sequence to allow the polymerizing agent to continue replication from the primers which are in stable duplex form with the target sequence, under the polymerizing conditions. The oligomers may contain a minimum of about 4 contiguous nucleotides which are complementary to a targeted *pilC* sequence; usually the oligomers will contain a minimum of about 8 contiguous nucleotides which are complementary to the targeted *pilC* sequence, and preferably will contain a minimum of about 14 contiguous nucleotides which are complementary to the targeted *pilC* sequence.

The oligomer, however, need not consist only of the sequence which is complementary to the targeted *pilC* sequence. It may contain in addition, nucleotide sequences or other moieties which are suitable for the purposes for which the oligomers are used. For example, if the oligomers are used as primers for the amplification of targeted *pilC* sequences via the

polymerase chain reaction (PCR), they may contain sequences which, when in duplex, form restriction enzyme sites which facilitate the cloning of the amplified sequences. Other types of moieties or sequences which are useful of which the oligomers may be comprised or coupled to, are those which are known in the art to be suitable for a variety of purposes, including the labeling of nucleotide probes.

In the basic nucleic acid hybridization assay, single-stranded analyte nucleic acid (either DNA or RNA) is hybridized to a nucleic acid probe, and resulting duplexes are detected. The probes for *pilC* sequences (natural or derived) are a length which allows the detection of these sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 20 nucleotides or more appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. For use as probes, complete complementarity is desirable, although it may be unnecessary as the length of the fragment is increased.

For use of such probes as agents to detect the presence of *pilC* sequences, the sample to be analyzed (which may be biological) may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. In order to form hybrid duplexes with the targeting sequence of the probe, the targeted region of the analyte nucleic acid must be in single-stranded form. The latter may occur naturally;

alternatively, it may be accomplished by denaturation. Denaturation can be accomplished by various techniques known in the art. Subsequent to denaturation, the analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte, and the resulting duplexes containing probe(s) are detected.

Detection of the resulting duplex, if any, is usually accomplished by the use of labeled probes; alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., doxetanes, particularly triggered dioxetands), enzymes, antibodies, and the like. Variations of this basic scheme are known in the art.

If the targeted *pilC* sequences are expected to be present at relatively low levels, amplification may be required for their detection. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT-tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT application 84/03520 and EPA124221 describe a DNA hybridization assay in which: (1) analyte is annealed to a single-stranded DNA probe that is complementary to an enzyme-labeled oligonucleotide; and (2) the resulting tailed duplex is hybridized to an enzyme-labeled

oligonucleotide. EPA 204510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands. A particularly desirable technique may first involve amplification of the target *pilC* sequences. The target *pilC* sequences in sera may be amplified, for example, to approximately 10^6 sequences/ml. This may be accomplished, for example, by the polymerase chain reactions (PCR) technique described by Saiki et al. (1986), by Mullis, U.S. Patent No. 4,683,195, and by Mullis et al. U.S. Patent No. 4,683,202. Amplification may be prior to, or preferably subsequent to purification of the *pilC* target sequence. For example, amplification may be utilized in conjunction with the assay methods described in U.S. Patent No. 4,868,105, or if even further amplification is desired, in conjunction with the hybridization system in EPO Publication No. 317,077.

Generally, in the PCR technique, short oligonucleotide primers are prepared which match opposite ends of a desired sequence. The sequence between the primers need not be known. A sample of polynucleotide is extracted and denatured, preferably by heat, and hybridized with oligomers which are oligonucleotide primers, which are present in molar excess. Polymerization is catalyzed by a template- and primer-dependent polymerase in the presence of deoxynucleotide triphosphates (dNTPs), and may also be in the presence of nucleotide analogs. This results in two "long products" which contain the respective primers at their 5'-termini, covalently linked to the newly synthesized complements of the original strands. The replicated DNA is again

denatured, hybridized with oligonucleotide primers, returned to polymerizing conditions, and a second cycle of replication is initiated. The second cycle provides the two original strands, the two long products from cycle 1, and two "short products" replicated from the long products. The short products contain sequences (sense or antisense) derived from the target sequence, flanked at the 5'- and 3'-termini with primer sequences. On each additional cycle, the number of short products is replicated exponentially. Thus, this process causes amplification of a specific target sequence.

It will be understood that "primer", as used herein, may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the target region to be amplified. Hence, a "primer" includes a collection of primer oligonucleotides containing sequences representing the possible variations in the sequence or includes nucleotides which allow a typical base pairing. One of the primer oligomers in this collection will be homologous with the end of the target sequence.

The amplified sequence(s) may then be detected using a hybridization assay which utilize nucleic acid multimers which bind to single-stranded analyte nucleic acid, and which also bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO 225,807, published June 16, 1987.

The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled

and the ingredients for labeling may be included in the kit in separate containers. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, buffers, as well as instructions for conducting the test using the kit ingredients.

The *pilC* DNA sequence information in the clones described in the Examples may be used to gain further information on the remaining sequence of the *pilC* gene from meningococci, for other possible alleles of *pilC* in *Neisseria*, as well as *pilC* in other relevant genuses and species. This information will aid in the characterization of the gene, and of its role in virulence of the pathogenic forms of microorganisms, including, for example, *Neisseria*, *Pseudomonas*, *Bacteroides*, *Moraxella*, and *Vibrio*. Moreover, this sequence information can lead to additional polynucleotide probes, polypeptides derived from *pilC*, multiple *pilC loci*, and antibodies directed against PilC epitopes which would be useful for the diagnosis and/or treatment of infections caused by the relevant pathogenic microorganisms.

The DNA sequence information in the above-mentioned clones is useful for the design of probes for the isolation of additional DNA sequences which are derived from as yet undefined regions of *pilC*. For example, labeled probes containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the ends of the DNA sequences shown in the Examples. These probes may be used to isolate overlapping DNA sequences within or adjacent to *pilC* from DNA libraries created from genomes of species having type 4 pilins. The resulting overlapping DNAs may then be used to

synthesize probes for identification of other overlapping fragments which do not necessarily overlap the DNAs whose sequences are given in the Examples. Thus, it is possible to sequence entire *pilC* genes utilizing the DNA sequences provided herein and the technique of isolation of overlapping DNAs derived from the *pilC* genes.

Methods for constructing DNA libraries are known in the art, and are discussed infra; for example, a method for the construction of *pilC* libraries in lambda-gt11 is discussed infra in Section IV.A. However, DNA libraries which are useful for screening with nucleic acid probes may also be constructed in other vectors known in the art, for example, lambda-gt10 (Huynh et al. (1985)). Another suitable vector for the creation of libraries may be EMBL3, which is a replacement vector which accepts inserts ranging from 9 to 23 kb in size. In general, methods for constructing DNA libraries is discussed in Maniatis et al, MOLECULAR CLONING, 2nd edition, (1989).

The sequence information derived from these overlapping *pilC* DNAs is useful for determining areas of homology and heterogeneity within the *pilC* gene(s), which could indicate the presence of different strains gonococci, meningococci, or other hitherto unrecognized pathogenic forms of *Neisseria*. It is also useful for the design of hybridization probes to detect PilC antigens or *pilC* nucleic acids in biological samples. Moreover, the overlapping DNAs may be used to create expression vectors for polypeptides derived from *pilC* gene(s).

The *pilC* DNA sequence information may also allow the construction of additional bacteriostatic agents for treatment of neisserial infections, in that they may block the expression of PilC and/or pilin assembly. For example, it may be used to derive

antisense polynucleotides. Antisense polynucleotides molecules are comprised of a complementary nucleotide sequence which allows them to hybridize specifically to designated regions of genomes or RNAs. Antisense polynucleotides may include, for example, molecules that will block protein translation by binding to mRNA, or may be molecules which prevent replication of DNA by replicase. They may also include molecules which carry agents (non-covalently attached or covalently bound) which cause the mRNA or genomic DNA to be inactive by causing, for example, scissions in these molecules. Antisense molecules which are to hybridize to *pilC* derived polynucleotides may be designed based upon the sequence information of the *pilC* DNA sequences provided herein, including those which would be isolated from additional DNA libraries. The antibacterial agents based upon anti-sense polynucleotides for *pilC* may be designed to bind with high specificity, to be of increased solubility, to be stable, and to have low toxicity. Hence, they may be delivered in specialized systems, for example, liposomes, or by gene therapy. In addition, they may include analogs, attached proteins, substituted or altered bonding between bases, etc.

Other types of drugs may be based upon polynucleotides which "mimic" important control regions of the *pilC* gene, and which may be therapeutic due to their interactions with key components of the system responsible for expression of the gene.

In addition to the specific methods described in the Examples, general methods are known which may be used in the practice of the invention. For example, general techniques used in extracting the genome from bacteria, including *Neisseria*, preparing and probing a DNA library, sequencing clones, constructing expression

vectors, transforming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for rowing cells in culture, and the like are known in the rt and laboratory manuals are available describing these techniques. However, as a general guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al. (1977)), the tryptophan (trp) promoter system (Goeddel et al. (1980)) and the lambda-derived P_L promoter and N gene ribosome binding site (Shimatake et al. (1981)) and the hybrid tac promoter (De Boer et al. (1983)) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with *E. coli*; if desired, other prokaryotic hosts such as strains of Bacillus or Pseudomonas may be used, with corresponding control sequences.

Eukaryotic hosts include yeast and mammalian cells in culture systems. *Saccharomyces cerevisiae* and *Saccharomyces carlsbergensis* are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983)), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968); Holland et al. (1978)), including the promoter for 3 phosphoglycerate kinase (Hitzeman (1980)). Terminators may also be included, such as those derived from the enolase gene (Holland (1981)). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are not naturally associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type

Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978)), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art.

Vectors suitable for replication in mammalian cells are known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding PilC epitopes into the host genome.

A vector which is used to express foreign DNA, and which may be used in vaccine preparation is Vaccinia virus. In this case the heterologous DNA is inserted into the Vaccinia genome. Techniques for the insertion of foreign DNA into the vaccinia virus genome are known in the art, and utilize, for example, homologous recombination. The insertion of the heterologous DNA is generally into a gene which is non-essential in nature, for example, the thymidine kinase gene (tk), which also provides a selectable marker. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (see, for example, Mackett et al. (1984), Chakrabarti et al. (1985); Moss (1987)). Expression of the polypeptide containing at least one immunoreactive PilC epitope then occurs in cells or individuals which are immunized with the live recombinant vaccinia virus.

Other systems for expression of desired polypeptides include insect cells and vectors suitable for use in these cells. These systems are known in the art, and include, for example, insect expression transfer vectors derived from the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcNPV), which is a helper-independent, viral expression vector. Expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive expression of heterologous genes. Currently the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed for improved expression. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; See Luckow and Summers (1989)).

Methods for the introduction of heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summer and Smith, Texas Agricultural Experiment Station Bulletin No. 1555; Ju et al. (1987); Smith et al. (1983); and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. The inserted sequences may be those which encode all or varying segments of the polyprotein, or other orfs which encode viral polypeptides. For example, the insert could encode the following numbers of amino acid segments from the polyprotein: amino acids 1-1078; amino acids 332-662; amino acids 406-662; amino acids 156-328, and amino acids 199-328.

The signals for posttranslational modifications, such as signal peptide cleavage, proteolytic cleavage, and phosphorylation, appear to be recognized by insect cells. The signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells. Examples of the signal sequences from vertebrate cells which are effective in invertebrate cells are known in the art, for example, the human interleukin 2 signal ($IL2_s$) which is a signal for transport out of the cell, is recognized and properly removed in insect cells.

Recombinant polynucleotides are inserted into host cells by transformation. Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972); Maniatis (1989)). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978). Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), or the various known modifications thereof. Other methods for the introduction of recombinant polynucleotides into cells, particularly into mammalian cells, which are known in the art include dextran-mediated transfection, calcium phosphate mediated transfection, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the polynucleotides into nuclei.

The recombinant polynucleotide may be in the form of a vector. Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 microgram of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 microliters buffer solution by incubation of 1-2 hr at 37° C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560.

Sticky ended cleavage fragments may be blunt ended using *E. coli* DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

Ligation mixtures are transformed into suitable cloning hosts, such as *E. coli*, and successful

transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

The desired recombinant DNA sequences may be synthesized by synthetic methods. Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984). If desired the synthetic strands may be labeled with ^{32}P by treatment with polynucleotide kinase in the presence of ^{32}P-ATP, using standard conditions for the reaction.

DNA sequences, including those isolated from DNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase using, as a primer, a synthetic oligonucleotide complementary to the portion of the DNA to be modified, and having the desired modification included in its own sequence. The resulting double stranded DNA is transformed into a phage supporting host bacterium. Cultures of the transformed bacteria, which contain replications of each strand of the phage, are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and conditions which permit hybridization with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

DNA libraries may be probed using the procedure of Grunstein and Hogness (1975). Briefly, in this

procedure, the DNA to be probed is immobilized on nitrocellulose filters, denatured, and prehybridized with a buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinyl pyrollidone, and Ficoll, 50 mM Na Phosphate (pH 6.5), 0.1% SDS, and 100 micrograms/ml carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42°C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'-^{32}P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

For routine vector constructions, ligation mixtures are transformed into *E. coli* strain HB101 or other suitable host, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared according to the method of Clewell et al. (1969), usually following chloramphenicol amplification (Clewell (1972)). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing. Sequencing may be by the dideoxy method of Sanger et al. (1977) as further

described by Messing et al. (1981), or by the method of Maxam et al. (1980). Problems with band compression, which are sometimes observed in GC rich regions, were overcome by use of T-deazoguanosine according to Barr et al. (1986).

An enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

Examples

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous

embodiments within the scope of the claims will be apparent to those of ordinary skill in the art.

Example 1

Isolation of PilC

An outer membrane preparation from *N. gonorrhoeae* strain $MS11_{ms}$ (P^+) contains small amounts of a 110 kd protein, PilC. This protein was enriched during alternate cycles of crystallization and solubilization of pili, unlike other outer membrane proteins that decreased in abundance by this procedure.

The materials and methods used for the isolation procedure were the following.

Bacterial strains and growth conditions

N. gonorrhoeae $MS11_{ma}$ (Meyer et al., 1984) and P^- and P^-n variants of $MS11_{mk}$ (Swanson et al., 1986) were kindly obtained from Dr. M. So and from Dr. M. Koorney, respectively. The gonococcal isolates UM01 and KH4318 have previously been described (Norlander et al., 1981). *N. gonorrhoeae* strains 605344 and 605103 were obtained from Dr. D. Danielsson, Örebro, Sweden, and strain 765 was isolated at the Department of Bacteriology in Umeå, Sweden. The commensal *Neisseria* species *N. lactamica* Nctc 10618 and *N. subflava* GN01 were obtained from Pharmacia, Uppsala, Sweden. These bacteria were grown at 37°C in a 5% CO_2 atmosphere on Difco GCB agar containing Kellogg's supplement. Piliated (P^+) and nonpiliated (P^-) variants were distinguished by colony morphology and passed as single colonies. *E. coli* strain Y 1090 (obtained from Promega Biotech) was used for plaque screening, DH5 (Hanahan, 1985) for molecular cloning, AA10 *recA* (Stoker et al. 1984) for isolation of minicells and TG1 (Gill et al., 1986) for propagation of M13 clones.

Preparation of pili and outer membranes

Pili were prepared essentially as described by Brinton et al. (1978). Gonococci (P^+Tr) from 80 GGB plates, grown for 18 h, were harvested in 0.05 M Tris-HCl pH 8.0 and 0.15 M NaCl, washed twice and resuspended in 40 ml 0.15 M ethanolamine pH 10.5. Pili were sheared off in a Sorvall Omnimixer, setting 3 for 30 s. The cell debris was pelleted at 13,000 g for 30 min at 4°C and the supernatant was dialyzed against 0.05 M Tris-HCl pH 8.0 and 0.15 M NaCl. The crystallized pili were pelleted at 13,000 g for 60 min., resuspended in 0.15 M ethanolamine pH 10.5, and centrifuged at 23,000 g for 60 min. The supernatant was dialyzed as described above against 0.05 M Tris-HCl pH 8.0 and 0.15 M NaCl. Several cycles of crystallization and solubilization were performed to produce pili preparations with high purity. Outer membranes of *N. gonorrhoeae* were prepared by the sarkosyl method described by Norquist et al. (1978).

Example 2

Preparation of Purified Anti-PilC Antibodies

The 110 kd protein present in purified $MS11_{ms}$ pili preparations was eluted from SDS polyacrylamide gels and rabbit antibodies were generated against the gel purified protein. The antiserum cross reacted extensively with the pilin protein in immunoblots and was therefore absorbed with extracts of *Pseudomonas putida* expressing the pilin subunit of *N. gonorrhoeae* on plasmid pGC02.

Pili preparations of *N. gonorrhoeae* $MS11_{ma}$ (P^+) crystallized 5 times were separated on 10% SDS-polyacrylamide gels using the buffer system of Laemmli (1970). These gels were stained in 0.25 M KCl and 1 mM

DTT for 5 min., the 110 kd protein band was sliced out, crushed and incubated in a buffer containing 0.05 M Tris-HCl pH 7.9, 0.1 mM EDTA, 5 mM DTT and 0.15 M NaCl at 4°C overnight. Gel pieces were removed by centrifugation prior to immunization of rabbits.

The achieved 110 kd-antiserum was extensively absorbed with *Pseudomonas putida* 2440 (Bagdasarien et al. (1983), carrying a recombinant plasmid, pGCO2, constructed as follows. The 1.0 kb *Hpa*I-*Eco*RI fragment of the pilus gene clone pNG1100 (Meyer et al. 1984) obtained from M. So was cloned into the *Hpa*I and *Eco*RI sites of pMMB66 (Fürste et al., 1986). The *pilE* gene is then under control of the *tac* promoter and induction with 1 mM IPTO resulted in high levels of pilin produced in *P. putida* 2440, but no extracellular pili structures were observed. Dense sonicated cultures of *P. putida* 2440/pGC02 were mixed in a 1:1 ratio with the crude antiserum. About 15 cycles of 1 h incubation and 30 min centrifugation at 25,000 g in the presence of 1 mM PMSF (phenylmethylsulfonylfluoride) at 4°C were performed.

The pili antiserum used in immunoblots was generated in a rabbit against highly purified pili preparations of *N. gonorrhoeae* $MS11_{ma}$.

In immunoblots 10 µg of boiled bacterial cells or the same amount of outer membranes were electrophoresed on 10% SDS-polyacrylamide gels. The proteins were transferred from the gel onto nitrocellulose sheets where their immunological cross-reaction with the 110 kd absorbed antiserum was tested using an immunoblotting protocol as described by Towbin et al. (1979).

Example 3

Specificity of Purified Anti-PilC Antibodies

The absorbed antiserum was used in immunoblots with whole cell extracts of a number of *N. gonorrhoeae* strains as well as commensal strains of Neisseria (Figure 2A). All strains of *N. gonorrhoeae*, except strain 605103, contained one or two high molecular weight protein species reacting with the antiserum. Strain 605103, unlike the other strains tested, was nonpiliated and no piliated variants could be obtained suggesting that it is a P^-n variant (Swanson et al., 1985). This was confirmed by Southern blot hybridization using an oligonucleotide probe corresponding to the 5' end of the *pilE* gene. No hybridization was obtained with this probe. The commensal *N. lactamica* Nctc10618, but not *N. subflava* GN01, contained a high molecular weight protein reacting with the 110 kd antiserum. Immunoblots against outer membrane preparations of P^+ and P^-n $MS11_{mk}$ showed the 110 kd protein to be present in the outer membrane in both of these MS11 variants.

Southern blot hybridization was accomplished as follows. Digested genomic DNA was separated on 0.7% agarose gels and transferred to nitrocellulose filters (Southern, 1975). After transfer and baking the filters were prehybridized in a mixture of 5 x SSC, 0.1% SDS, 5 mM EDTA, 5 x Denhardt's solution and 100 μg/ml of sonicated calf thymus DNA at 65°C for 2-6 h. ^{32}P-labeled probe (multiprime DNA labelling system, Amersham International) was added and hybridization was performed for 12-15 h at the same temperature. The filters were washed in 2 x SSC with 0.1% SDS and in 0.2 x SSC with 0.1% SDS for 2 x 15 min each, dried and exposed to Kodak XRP film at -80°C.

A 21-base-long oligonucleotide complementary to the signal peptide coding region of *pilE* (5'-GCCTTTTTGAAGGGTATTCAT-3') was ^{32}P-labeled with T4 polynucleotide kinase and used to probe *Cla*I-digested genomic DNA. The blot was prehybridized at 37°C in a mixture containing 2 x Denhardt's, 0.1% SDS, 2.5 mM EDTA, 5 x SSC and 100 μg/ml sonicated calf thymus DNA, hybridized at 37°C and washed in 2 x SSC for 5 min. MS11$_{mk}$(P$^+$) gave a 4 kb hybridization fragment, whereas MS11$_{mk}$(P$^-$n) and 605103 gave no hybridization signal.

Example 4

Molecular cloning of the *pilC1* gene encoding a 110 kd protein

Chromosomal DNA from *N. gonorrhoeae* MS11$_{mk}$(P$^+$) was used to construct a λgt11 library. The library was screened with the absorbed 110 kd antiserum and one positive clone out of 10,000 plaques was found, containing an 800 bp insert. A lysogen of this positive λgt11 clone was examined in immunoblots and a fusion protein with an estimated size of 150 kd reacted with the antiserum (data not shown). The 800 bp insert was purified, labeled with ^{32}P, and used as a probe to screen a plasmid library from *N. gonorrhoeae* MS11$_{ms}$. Six clones out of 10,000 hybridized with the probe. Restriction maps for these partially overlapping six clones are shown in Figure 1.

In Figure 1, plasmids pABJ04-09, which all belong to locus 1, were isolated from a plasmid library using the 800 bp insert from λgt11 as a probe. The λgt11 insert (from locus 2) has an additional *Sal*I site not found in the plasmid clones. The position of the *pilC1* gene and direction of its transcription (indicated by an arrow) were determined in *E. coli* minicells. Three

thickened lines, with an arrow-head at each end, indicate fragments used as probes in Southern hybridizations. I.e., the 800 bp insert from λgt11, the *Eco*RV$_1$-*Eco*RV$_2$ (1.3kb) and the *Eco*RV$_3$-*Hin*dIII$_4$ (0.8 kb) fragments of pAGJ04. Triangles mark the location of two mTnCm insertions in pABJ04. The resulting plasmids, pABJ)4::mTnCm-12 and pABJ04::mTnCm-14 were used, were used to inactivate *pilC1* and *pilC2*.

The six plasmid clones, pABJ04-09, were transformed into the minicell producing strain AA10 to monitor expression of plasmid encoded [^{35}S]methionine labeled proteins. The *E. coli* minicell strain AA10 was transformed with plasmid DNA (pABJ04-09) and chromosome deficient minicells from these strains were purified over sucrose gradients (Thompson and Achtman, 1978). The plasmid-encoded proteins were labeled in the presence of 80 μCi [^{35}S]methionine in minimal salts medium and 1% methionine assay medium (Difco). After lysis of the minicells in sample buffer (Laemmli, 1970) the proteins were electrophoresed on an SDS-polyacrylamide gel, the gel was dried and exposed to X-ray film (Kodak X-OmatAR).

Plasmid pABJ04 expressed minute amounts of three high molecular weight proteins, 113, 111 and 108 kd in size, as well as a number of lower molecular weight protein species not produced from the vector control. The three high molecular weight bands were missing in pABJ05 and pABJ06 but three novel lower molecular weight protein species had appeared, suggesting that pABJ05 and pABJ06 are deleted for the 3' end of a gene, denoted *pilC1*, and that this gene is responsible for all three high molecular weight species. This suggested that the distal end of the gene must be located between the *Mlu*I$_1$ and *Mul*I$_2$ sites (Figure 1). The observation that plasmid pABJ07 did not express any high molecular proteins

tentatively located the 5′ end of the gene to a region 0.5-1.2 kb to the right of the *Eco*RV$_3$ site. The size for a gene encoding a 110 kd protein is ~3 kb which is in agreement with these mapping data.

Example 5

Identification of a Second Gene Encoding PilC

The 800 bp insert in λgt11 contains a single *Sal*I site not present in the region on pABJ04 which hybridized to this fragment, suggesting that there is more than one *pilC* locus in the genome of *N. gonorrhoeae* MS11. This was confirmed in Southern blot hybridizations in which three different *pilC* fragments were used to probe *Sma*I and *Cla*I digested genomic DNA. The 800 bp fragment from λgt11 hybridized in a Southern blot to two *Cla*I (18 and 8 kb) and *Sma*I (13 and 4.5 kb) fragments of DNA prepared from *N. gonorrhoeae* MS11$_{mk}$. Since the probe does not contain any internal *Cla*I or *Sma*I sites, there are presumably two copies of the 3′ end of *pilC* in the MS11 genome. The 1.3 kb *Eco*RV$_1$ - *Eco*RV$_2$ fragment of pABJ04 carries the central region of *pilC1*. This probe hybridized to the same two *Cla*I fragments and to four *Sma*I fragments, two of which are the same size as the two *Sma*I fragments identified with the 800 bp probe (13 kb and 4.5 kb). Hybridization with the 800 bp probe was more extensive to the 8 kb *Cla*I and the 4 kb *Sma*I fragment whereas the reverse was found with the 1.3 kb *Eco*RV$_1$ - *Eco*RV$_2$ fragment from pABJ04 strongly suggesting that the two genomic copies of *pilC* show a significant sequence variation in the 3′ as well as in the central region. A probe corresponding to the 5′ region of *pilC1* was also used in Southern hybridization experiments. This 0.8 kb *Hin*dIII$_4$ - *Eco*RV$_3$ fragment hybridized to two *Cla*I (18 kb and 4 kb) and *Sma*I (25 kb and 7 kb) fragments

with seemingly equal efficiency. The hybridization pattern was identical using DNA from $MSII_{ms}$. Taken together these hybridization data indicate that *N. gonorrhoeae* MS11 contains two complete copies of *pilC*. Furthermore the two genes appear to be more homologous in their 5' as compared to their central and 3' regions.

The results indicate that the 800 bp insert from λgt11 carries information from *pilC2* whereas the clones pABJ04-09 must carry information from *pilC1*. Finally *pilC2* must be located >2 kb from either end of *pilC1*. The DNA sequence of the 3'-end of the *pilC2* fragment is shown in Figure 7. The sequence showing the putative amino acids encoded therein are shown in Figure 8. A comparison of the analogous portions of *pilC2* (top) and *pilC1* (bottom) DNA sequences, and the putative amino acids encoded therein are shown in Figure 9.

The 800 bp fragment from *pilC2* was also used to probe digested genomic DNA from *N. gonorrhoeae* strains UM01, 765 and 605103. The latter isolate does not express detectable levels of the 110 kd protein. Strain UM01, unlike MS11, contained only one *Cla*I fragment of 15 kb that hybridized to the probe (data not shown). Hence, this strain may contain only one copy of *pilC*. Strain 605103 and 765, on the other hand, each seem to contain two copies of *pilC* since two *Cla*I and two *Sma*I fragments hybridized to the 800 bp probe.

The commensal *N. lactamica* Nctc10618 DNA digested with *Cla*I and *Sma*I also hybridized with the 800 bp probe. Since only one band hybridized in each case this strain may contain only one copy of *pilC*. In contrast, *N. subflava* GN01 did not hybridize to the 800 bp *pilC2* probe using the same stringency.

Example 6

Characterization of the *pilC* Genes

The pilC1 gene on pABJ04 is translationally out of frame

The amino terminal sequence of gel purified 110 kd protein from strain $MS11_{ms}$ (P^+) was determined by sequential Edman degradation. For aminoterminal sequence determination automated Edman degradations (Edman and Bregg, 1967) were performed in an updated Beckman 890C spinning cup sequencing sequencer. The sequencing procedure and the method for analysis of the 3-phenyl-2-thiohydantoin derivatives been described (Engström et al., 1984). Considerable difficulties were encountered in the method probably due to blocking of the N-terminus. As a result, only the residues from position 4 to 10 were obtained (Figure 2).

The 3.3 kb *Hind*III_4 - *Mlu*I_1 fragment encompassing the entire *pilC1* gene was sequenced on both strands using the dideoxy sequencing method adapted for single stranded DNA.

Purified DNA fragments from pABJ04 and PCR-amplified 5' end of *pilC1* was subcloned into M13 vectors (Sanger et al., 1980: Yanish-Perroa et al., 1985) and sequenced using the chain termination method of Sanger et al. (1977). Primers used were the M13 17-mer universal primer and oligonucleotides synthesized at Symbicon, Umeå, Sweden or at the Department of Biochemistry, Washington University, St., Louis, MO, USA.

The results of the sequencing showed that the *pilC1* contained one single open reading frame of 997 codons (from left to right in Figure 2) and starting at an AUG codon 195 bp from the *Hind*III_4 site. Codons 7-12 in this open reading frame corresponded to amino acids 5-10 in the sequence of the gel purified protein. The AUG

codon in the beginning of the long open reading frame was not preceded by a typical Shine-Dalgarno sequence. Moreover, since the 110 kd protein is located in the outer membrane of *N. gonorrhoeae*, we expected the protein to be translated with a signal sequence. When examining the nucleotide sequence, an AUG codon was found in frame 1 that was preceded by a typical Shine-Dalgarno sequence (-AGGAA-). The sequence following this AUG codon would encode a typical signal peptide with basic amino acids in the amino terminal region and a hydrophobic central region. However, no signal peptidase cleavage site could be predicted following the rules of von Heijne (1983). A tract of 12G residues was found in the region encoding the putative signal peptide for PilC. Addition of one G residue or the loss of two would align the long open reading frame with the AUG codon in frame 1. The translated region in frame 2 contains a putative signal peptidase cleavage site between Ala and Gln. A cleavage at this site would align the determined amino acid sequence at positions 5-10 for the 110 kd protein with the deduced amino acid sequence. The data therefore suggested that the cloned *pilC1* gene is out of frame due to frameshifting in the region encoding the signal peptide.

Figure 2 shows the nucleotide sequence and the deduced amino acid sequence of the 5'-end of *pilC*. The amino-terminal sequence of gel purified PilC from $MS11_{ms}(P^+)$ is shown in a box below frame 2, a 997 amino acid long open reading frame that would code for a protein about 110 kd in size. Frame 1 contains 41 amino acids and is preceded by a putative Shine-Dalgarno sequence (underlined). Two horizontal lines mark a stretch of 12 G residues. An addition of one G in this region would align the ATG (boxed) in frame 1 with frame

2. Numbers above the sequence show base positions relative to the *Hind*III$_4$ site (=0) located on pABJ04. The position of two 24 bp oligonucleotide primers (opposite stands) used for PCR amplification, are indicated above the sequence by hatched bars.

Figure 3 shows the nucleotide sequence of the sense strand of the *pilC1* gene.

Figure 4 shows the shows the nucleotide sequence of the sense strand of the *pilC1* gene and the amino acids encoded therein.

Genetic inactivation of *pilC2* but not *pilC1* abolishes expression of the 110 kd protein in MS11

Plasmid pABJ04 was mutagenized in *E. coli* by a transposon mini-Tn3 derivative, mTnCm. The shuttle mutagenesis system developed by Seifert et al., (1986) using a miniTn3 carrying the chloramphenicol resistance gene was kindly provided by Dr. M. So. Mutagenesis of pABJ04 with mTnCm and transformation of *N. gonorrhoeae* were performed as previously described (Seifert et al., 1990). MiniTnCm insertions at 30 different positions in pABJ04 were identified, two of which mapped within the PilC gene. Piliated *N. gonorrhoeae* MS11$_{mk}$ were transformed with 2 μg plasmid DNA, transformants were selected for on plates containing 10 μg/ml chloramphenicol for the single mutants and 30 μg/ml chloramphenicol for the double mutants.

Only two mTnCm insertions had occurred in *pilC1* (Figure 1). Truncated protein species were seen in minicells with the mTnCm-14 insertion located 0.5 kb from the 3' end of *pilC1* but not with the mTnCm-12 insertion located 0.5 kb from the 5' end of the gene. Both insertion mutants were used in a gene replacement experiment. Plasmids pABJ04::mTnCm-12 and pABJ04::mTnCm-

14 were linearized with BamHI and transformed into *N. gonorrhoeae* $MS11_{mk}$ (P^+) and transformants resistant to 10 μg/ml of chloramphenicol were selected. Forty-eight P^+ transformants (24 from each experiment) were assayed for the presence of PilC in immunoblots. All these transformants remained capable of expressing the PilC protein. Genomic DNA was prepared from seven of the chloramphenicol resistant transformants (five from pABJ04::mTnCm-12 and two from pABJ04::mTnCm-14), cleaved with ClaI and PvuII and used in Southern blot experiments using the $EcoRV_1$ - $EcoRV_2$ fragment of pABJ04 as a probe. The 8 kb ClaI fragment was unaffected in the mutants whereas the 18 kb ClaI fragment had been replaced by a 20 kb fragment. PvuII cleaves within the 1.6 kb mTnCm element. The probe detected an 8 kb PvuII fragment in both parent and mutant DNA. In the mutants, a novel PvuII fragment appeared that was 6.2 kb in size in five transformants obtained with pABJ04::mTnCm-12 and 4.8 kb in size in two transformants with pABJ04::mTnCm-14. To confirm the insertion of mTnCm, a 250 bp EcoRI-HindIII fragment of the CAT GenBlock (Pharmacia, Sweden), containing the PvuII site, was used as a probe. It detected the larger of the two ClaI fragments as well as the 6.2 kb PvuII fragment. In addition, a 2 kb PvuII fragment not covered with the *pilC* probe was detected. These data demonstrate that we have obtained gene replacements in *pilC1*, whereas *pilC2* was unaffected in all seven P^+, $PilC^+$ transformants. A rapid hybridization was done to screen the remaining 41 P^+ transformants. All but one had mTnCm inserted in *pilC1*. The remaining transformant had an intact locus 1 and 2 and must therefore contain mTnCm elsewhere in the gonococcal chromosome.

In the same transformation experiments, the frequency of P^- colony variants was about five-fold higher as compared with that occurring normally in strain $MS11_{mk}(P^+)$. Two P^-mTnCm-12 transformants isolated at 10 μg/ml of chloramphenicol were also analyzed by Southern blot hybridization using the $EcoRV_1$ - $EcoRv_2$ fragment of pABJ04 and the EcoRI - HindIII fragment of the CAT GenBlock. Each of these mutants carried mTnCm in *pilC2* as evidenced by a replacement of the 8 kb ClaI fragment by a fragment 9.5 kb in size that hybridizes to both probes. These *pilC2*::mTnCm insertion mutants did not express PilC as determined by immunoblot analysis.

A P^+, *pilC1*::mTnCm-12 mutant was retransformed with DNA prepared from a P^-, *pilC2*::mTnCm-12 mutant and colonies growing at 30 μg/ml of chloramphenicol were selected to obtain double mutants in *pilC*. All resistant transformants were P^-, and when analyzed by Southern blot hybridization all contained mTnCm in both *pilC1* and *pilC2*. Electron microscopy revealed that the P^+, *pilC1*::mTnCm-12 mutant still expressed pili albeit at a slightly lower level than the $MS11_{mk}(P^+)$ parental clone, whereas the P^-, *pilC2*::mTnCm-12 was completely bald as was the *pilC1*, *pilC2* double mutant.

Immunoblot analyses were performed on the P^+ *pilC1*::mTnCm-12 mutant, the P^-, *pilC2*::mTnCm-12 mutant and the P^-, *pilC1*::mTnCm-12, *pilC2*::mTnCm-12 double mutant, using PilC and pili antisera. Inactivation of *pilC1* did not abolish expression of PilC or the pilin. Inactivation of *pilC2* totally abolished expression of PilC but did not affect expression of pilin. The *pilC1*, *pilC2* double mutant was $PilC^-$ but produced only low levels of pilin. Taken together these data imply that *pilC2* but not *pilC1* is expressing PilC in the MS11

variant under study. Moreover, inactivation of *pilC2* but not *pilC1* was associated with a loss of piliation.

P^+ revertants occurred spontaneously at a low frequency in the *pilC2*::mTnCm-12 mutants. These revertants expressed pili as determined by electron microscopy and also expressed PilC. It is likely that PilC expression is due to in-frame switching in *pilC1*.

Figure 5 shows the nucleotide sequence of the sense strand of the *pilC1* gene, and the effect of frame shift on the putative gene products encoded therein.

The *pilC* genes of *N. gonorrhoeae* vary in the length of the G tract

Polymerase chain reaction (PCR) with Taq polymerase was used to analyze the 5' region of *pilC* using two 24 base long synthetic oligonucleotides based on the sequence of *pilC1* (Figure 2). These oligonucleotides would generate an amplified fragment of 149 bases as judged from the sequence obtained from pABJ04.

Polymerase chain reaction was carried out in 100 µl containing 50 ng of genomic DNA or 5 ng of plasmid DNA. 1.0 µM of each oligonucleotide, 200 µM of each nucleotide, 0.001% gelatin, 1.5 mM $MgCl_2$, 10 mM Tris (pH 8.3), 50 mM KCl, 0.25 µl 1 mCi/ml [^{32}P]dATP and 2 U of Taq Polymerase (Perkin Elmer Cetus). The samples were passed through 25 cycles: 2 min at 50°C, 1 min at 94°C and 3 min at 72°C in a Thermal Cycler (Perkin Elmer Cetus). Aliquots of the DNA fragments were denatured at 95°C for 2 min and electrophoresed on standard denaturing sequencing gels.

The amplified products from $MS11_{mk}$(P^+) DNA were 149 and 150 long respectively. In addition, two less abundant products of 151 and 148 bases were seen. The

amplified products were electroeluted and cloned into M13mp18, and twenty phage clones were sequenced using a universal primer. Four different sequences were obtained (Figure 6).

Figure 6 shows the nucleotide sequence of PCR amplified fragments demonstrating a variation in the length of the G tract and sequence differences in the 5'-region of the *pilC* genes. The two oligonucleotide primers used for the PCR are shown in Figure 2. Amplified DNA was cloned into M13mp8 and sequenced. Shown are the complete nucleotide sequence in between the two primers. In-frame sequences are translated and the G stretches are underlined. The putative cleavage sites are marked with arrows. Genomic DNA from *N. gonorrheae* strains MS11(P^+.Pilc$^+$), UM01(P^+.Pilc$^+$), 765 (P^+.Pilc$^+$) and 605103 (P^-n.Pilc$^-$), and purified DNA from pABJ04/AA10(*rec*A) was used in the PCR.

Variant patterns 1a and 1b were identical to each other and to the cloned sequence on pABJ04 except for the presence of 11 instead of 12 G residues in the G tract of 1b. The G tract of sequence 2a was 13 residues long indicating that the sequence is in frame. In addition, this sequence differed from *pilC1* by four basepair substitutions outside the G tract, including an AAA lysine codon four triplets downstream of the putative signal peptide processing site which is in agreement with the lysine residue found in the fourth position of the gel purified 110 kd PilC protein. Sequences 1a and 1b contained CAA, the codon for Gln, at the same position. Sequence 2b was identical to 2a except for the presence of 12 G residues in the G tract. These data are compatible with sequence 1 being from *pilC1* and sequence 2 from *pilC2* and further support that *pilC2* must be the expressed locus in the MS11(P^+) variant we are studying.

Strain UM01 apparently only contains one copy of *pilC*. DNA from this strain generated five amplified fragments ranging in size from 148 to 152 bp in the PCR reaction. The most abundant fragments were 149-151 bp long. Among ten M13 clones, three variant 1 sequences were found (a,b,c) that differed only in the number of G residues (11-13) in the G tract (Figure 6) supporting the hybridization data that this strain contains only one *pilC* gene. Since a PilC protein is expressed from UM01 we suggest that the majority of cells has 13 Gs in the G tract.

Strain 765 contains two *pilC* loci, both of which seem to be translationally ON based on the presence of two high molecular weight proteins reacting with the absorbed PilC antiserum. A number of amplified fragments were seen after the PCR reaction ranging in size from 149 to 153 bases. Three variant sequences were found among nine clones (Figure 6). The G tract was 13 residues long in variant 3a (in frame) and 14 (out of frame) in variant 3b whereas sequence variant 4 contained 11 G residues in the G tract. Variant sequence 4 contained four additional nucleotides (-CAGG-) distal to the G tract relative to variant sequences 1, 2 and 3, indicating that the amplified product with 11 Gs from this variant sequence is 152 long and out of frame. Two PCR amplified products 152 and 153 in length were obtained from strain 765 suggesting that in frame variants of sequence 4 might be present in the DNA prepared from this strain.

Strain 605103 carries two *pilC* copies, both of which seem to be translationally OFF. The amplified fragments were 148 and 149 bases in size. Out of eight M13 clones only variant 1a and 1b sequences were found, with 11 and 12 Gs in the G tract respectively. Consequently, we were unable to find an in frame sequence

variant from this strain. We do not know if the 5′ ends of the two *pilC* genes are identical in this strain or if one *pilC* gene differed from *pilC1* in the region corresponding to the oligonucleotides used for amplification. In the latter case we would not expect to obtain any amplified products from the second copy.

The only in frame variant found in DNA amplified from *N. gonorrhoeae* carried 13 Gs in the G tract. To see if variants with 10Gs arise in products expressed in *E. coli*, PCR amplified products were generated from pABJ04 purified from *E. coli* strain AA10, using the same two oligonucleotide primers as before. Out of 12 sequenced clones, two carried 10 Gs in the G tract (Figure 6). The majority of clones (seven) carried 12 Gs as expected. It is likely that the PCR amplification products are not representative of the original DNA population. However, the distribution of variation in the G tract is consistent with a model in which only one G residue is gained or lost at one given event. Since AA10 is *recA*, frameshift mutations in the G tract in *E. coli* occur independently of the RecA protein.

N. meningitidis contains two *pilC* loci

Southern blot hybridizations using MS11 *pilC1*⁻ specific probes identified multiple fragments when meningococcal genomic DNA is digested with a variety of restriction endonucleases. PCR amplification using two 24-base oligonucleotides from the 5′ end of MS11 *pilC1* as primers yields multiple fragments ranging in size from 148 to 151 bases. DNA sequencing of fragments cloned into phage M13 identifies two classes of sequences, as in *N. gonorrhoeae*, which differ outside the G-tract. Variation occurred within each class with respect to the number of G's in the G-tract. Therefore, *N. meningitidis* must

carry two *pilC* loci, as does *N. gonorrhoeae*, which should also be under the control of translational frame shifting.

Both *pilC* loci are cloned from *N. meningitidis* by generating an EMBL3 library and screening this library with *pilC1*⁻ and *pilC2*⁻ specific DNA from *N. gonorrhoeae* strain MS11.

Genomic DNA from *N. meningitidis* is partially digested with Sau3A and fragments ranging from 9 to 20 kB are ligated into the lambda EMBL3 vector. Because of the packaging constraints of the phage, only those lambdas which contain DNA fragments of this size will be packaged (i.e., are viable). The library thus constructed can be screened with genomic oligonucleotide or cloned gene probes following selection in a lysogenic *E. coli* strain. (Frischart, A.M. et al (1983), J. Mol. Biol. 170:827). Preferably, full length clones are identified by screening for clones hybridizing to both the 5′ and 3′ ends of *pilC*. If full length clones cannot be obtained from the EMBL3 library, *pilC* specific probes may be used to screen a plasmid library from the same strains.

Translational fusion proteins with β-galactosidase may also be screened for in a λgt11 library, using β-galactosidase and PilC⁻ specific antisera in Western immunoblots. β-galactosidase-PilC1 and β-galactosidase-PilC2 fusion proteins are purified from the cytoplasm of recombinant *E. coli* and used to raise specific antisera.

Example 7

The immunobiological properties of PilC

PilC is located in the outer membrane of *Neisseria*. The immune response during natural infection can be assessed by screening convalescent sera for anti-

PilC antibody. The presence of only two *pilC* loci suggests that PilC is only moderately variable, however. This together with its essential role in pilus biogenesis makes PilC attractive as a potential vaccine candidate.

Two types of PilC translational fusions using alkaline phosphatase and β-galactosidase are generated. In the first instance a secreted fusion protein is obtained that may associate with the outer membrane. In the second instance the fusion proteins may accumulate in the cytoplasm as inclusion bodies. The construction schemes for such fusion proteins uses techniques known in the art. Tn*phoA* insertions on plasmid pABJ04 in *E. coli* are generated, and a $PhoA^+$ phenotype is screened for as blue colonies on media containing the chromogenic substrate XP. If such clones have the *phoA* gene in frame with an in frame variant of *pilC1* the fusion product should be able to cross the cytoplasmic membrane where it can be analyzed by Western immunoblots using an alkaline phosphatase specific antiserum and our PilC antiserum raised against gel-purified PilC2 from MS11(P^+). *LacZ::pilC* fusions are generated by cloning different segments of *pilC* into a *lacZ*$^-$ containing vector used to generate translational fusions. Similar constructs are performed on each of the two *pilC* genes from *N. meningitidis*. Antisera are generated against fusion proteins after their purification using conventional protocols. These antisera are extensively adsorbed with extracts of *E. coli* expressing alkaline phosphatase and β-galactosidase, and used in Western immunoblots and ELISA assays against a panel of *Neisseria gonorrhoeae* and *Neisseria meningitidis* strains. Antisera raised against fusion proteins carrying the major portion of PilC are also analyzed in Western blots using *E. coli* expressing fusion proteins containing only smaller regions of PilC.

The results of these studies should show which regions in PilC are immunodominant.

The *pilC1* and *pilC2* genes are highly homologous in their 5' ends whereas the homology is considerably less pronounced in the central and 3' region.

In addition, the entire *pilC2* gene from *N. gonorrhoeae* MS11(P^+) is cloned and sequenced. Algorithms are used to search for potential T-cell epitopes (amphipathic helical conformation) and β-cell epitopes. Polypeptides containing the predicted epitopes are tested to determine if they can prime mice for an enhanced immune response to PilC1 and PilC2.

Specific PilC antisera are used in immunoelectromicroscopy with piliated *Neisseria* cells as well as with purified pili to see if PilC is physically connected with the pilus fiber.

Neisseria is grown in the presence of different dilutions of $PilC^-$ specific antibodies. Bactericidal effect exerted by the antiserum, effects on piliation, and effects on bacterial attachment to corneal primary culture cells are monitored. Binding assays to epithelial cells are described in Tjia, K.F. et al. (1988), Graefe's Arch. Clin. Exp. Opthalmol. 226:341-345.

Example 8

Identification and characterization of genes located adjacent to *pilC*

The *pilC1* and *pilC2* loci are part of a larger duplication that extends both 5'- and 3'- of *pilC*. We know from our work with *E. coli* that strains may contain multiple gene clusters for the same class of pili. In one case we have shown that the only difference between two duplicated

gene clusters (*pap* and *prs*) resides in the adhesion genes such that each cluster gives rise to serologically identical pili binding to different cell surface receptors.

mTn*cm* mutagenesis in the region upstream and downstream of *pilC1* is performed to generate allelic replacements in the *pilC1* and *pilC2* regions on the chromosome. Since the two regions are highly homologous we expect to obtain for each insertion allelic replacements in either region. Double mutants are generated as before by isolating DNA from mutants carrying insertions in the *pilC2* region transforming P^+ variants carrying the same insertion in the *pilC1* region and select for transformants resistant to 30μg/ml of chloramphenicol. These double mutants are examined for piliation, pilins expression, and binding to corneal primary culture cells.

Example 9

Phase variation in gonococcal pili expression can be caused by frameshift mutations in *pilC*

If PilC is required for pilus formation, we would expect some P^- progeny arising from a P^+ clone to accumulate unassembled pilin in the absence of PilC. Nonpiliated (P^-) colonies were derived from $MS11_{mk}$(P^+), restreaked, and tested for the presence of PilC and pilin in immunoblots with the PilC and pili antisera. Five out of eight P^- clones did not produce detectable levels of PilC, but expressed the pilin subunit. The remaining three P^- clones expressed PilC but not pilin. The molecular mass of the pilin subunit was the same in the P^-, $PilC^-$ variants as in $MS11_{mk}$(P^+, $PilC^+$). However, the former in addition produced a protein reacting with the pili antiserum that was 16 kd in size. Since $MS11_{mk}$ only

contains one expression site for pilin we believe that this protein species represents a proteolytic degradation product of the pilin and may be identical to the S-pilin previously described (Haas et al., 1987). Several independent P^- clones were isolated from one P^-, $PilC^-$ clone. They all remained $PilC^-$ and retained expression of pilin. Piliated (P^+) revertants were also obtained from the same P^-, $PilC^-$ clone. These P^+ revertants occurred at about a tenfold lower frequency (10^{-4}) than P^- derivatives from a P^+ clone. All P^+ revertants from a P^-, $PilC^-$ clone had regained expression of PilC. All but one expressed a pilin with the same molecular weight as the nonpiliated parent. However, the low molecular weight pilin degradation product was much less abundant in the P^+, $PilC^-$ revertants. It was possible to obtain P^+ revertants from other P^-, $PilC^-$ clones as well, all of which expressed PilC.

The *pilE* gene from one set of PilC switches was PCR amplified and sequenced directly. The P^-, $PilC^-$, $pilin^+$ variant 8 carried eight amino acid changes in the pilin relative to the parental clone $MS11_{mk}$. The pilin sequence of the P^+, $PilC^-$ backswitcher 8:1 was identical to variant 8. Thus, the backswitching from P^- to P^+ colony morphology was not associated with any alteration in the pilus subunit protein implying that the change in colonial morphology was due to the switch in PilC expression.

Strain $MS11_{mk}$(P^-), variants 8(P^-) and 8:1 (P^+) were also examined by transmission electron microscopy. Electron microscopy was performed with a JEOL 100CX microscope with 200-mesh copper grids coated with thin films of 2% Formvar. The bacterial colonies were carefully overlaid with buffer [10 μM Tris-HCl (pH 7.5), 10 μM magnesium chloride] and the cells were allowed to

sediment for 15 min on a grid. The grids were washed with water, negatively stained with 1% sodium silicotungstate (pH 7.0) and then washed again.

The electron micrographs showed that the MS11(P^+) parental cells were heavily piliated and pili were often seen to aggregate. In contrast most cells of variant 8(P^-) were nonpiliated. One or two pili were found on ~10% of these cells. All cells of variant 8:1 (P^+) were piliated, carrying -10-40 fibers/cells. No aggregation of individual fibers was seen. These data confirm that the observed changes in colonial morphology reflect alterations in expression of pili. Therefore, phase variation of gonococcal pili may not only be caused by recombination events occurring in the *pilE* locus (Bergstrom et al, 1986; Swanson et al., 1986) but also by frameshift mutations in *pilC*.

Example 10

Immunogenicity of PilC

In order to predict a region of PilC which would have a high probability for antigenicity, residues 300 to 700 of the putative PilC1 protein encoded within *pilC1* were analyzed for antigen index, hydrophilicity, and hydrophobicity using standard computer-modelling methods. The analysis indicated that the PilC1 polypeptide fragment containing residues 300 to 700 would have several regions with a high antigen index, high hydrophilicity, and a high likelihood for location in an external domain.

The immunogenicity of a recombinant polypeptide expressed from the DNA encoding amino acids 300 to 700 was examined. The region of DNA encoding amino acid residues 300 to 700 was amplified by polymerase chain

reaction (PCR) using the following oligonucleotide primers.

5' GGC TAG GTG GCA TAT GAA AGA TAC CGG 3'

and

5' TTT GCA ATC GGG GAT CCT* C*A*G GTG TCT TTC 3'

These primers incorporate an NdeI and a BamHI restriction endonuclease site (indicated by the underlined nucleotides), respectively. A termination codon (indicated by the asterisks) was also incorporated. The PCR amplified DNA was then ligated into the vector pET3a (between the NdeI and BamHI sites). The recombinant vector pET3a is used in the inducible expression system described by Studier et al. (1990), using the protocol described therein. Strain BL21 (DE3) was transformed with the pET3a-*pilC* (300-700) vector, and the transformed strain used for the expression of the PilC (300-700) peptide.

The expression products after induction were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (PAGE) using an 8% acrylamide gel and the standard 25 mM Tris base / 250 mM glycine / pH 8.3 / 0.1% sodium dodecyl sulfate (SDS) electrophoresis buffer. After electrophoresis, the gel was fixed and then stained with Coomassie Blue according to standard protocols, and the production of PilC(300-700) was confirmed by the detection of the presence in the gels of an abundant, appropriately-sized peptide of approximately 46kD.

In order to detect the immunogenicity of the PilC(300-700) product, the region of the SDS-PAGE gel containing the PilC(300-700) polypeptide was excised from parallel unstained lanes, homogenized, and the protein eluted into a buffer of 0.1% Triton X-100 in water by

passive diffusion. Rabbits were given a priming intradermal injection of homogenized gel slices (containing approximately 500μg PilC(300-700) protein), followed 3 weeks later by a subcutaneous boost (of approximately 500 μg of eluted protein). An initial test serum was then collected after an additional 14 days. All of the test animals yielded a specific high titer antibody response. The antibodies induced by PilC(300-700) were immunologically reactive not only with that polypeptide (i.e., PilC1(300-700)), but also with native PilC1 and native PilC2.

The results demonstrate, inter alia, the following. PilC contains antigenic epitopes that can elicit a strong immunogenic response. At least some of the immunogenic epitopes are shared (cross-reactive) between PilC1 and PilC2, despite differences in primary amino acid sequence. The technique of subcloning discrete portions of the PilC protein under control of an inducible promoter allows mapping of antigenic epitopes. Sufficient quantities of specific oligopeptides of known antigenicity can be produced for use in screening the *in vivo* immune response after exposure to the intact pathogen.

The following listed materials are on deposit under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

Description	ATCC No.	Deposit Date
pABJ03 in *E. coli* (DH5)	68519	Jan. 28, 1991
pABJ04 in *E. coli* (DH5)	68520	Jan. 28, 1991

Upon allowance and issuance of this application as a United States Patent, all restriction on availability of these deposits will be irrevocably removed; and access to the designated deposits will be available during pendency of the above-named application to one determined by the Commissioner to be entitled thereto under 37 CFR 1.14 and 35 USC 1.22. Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. The deposited materials mentioned herein are intended for convenience only, and are not required to practice the present invention in view of the descriptions herein, and in addition these materials are incorporated herein by reference.

Industrial Applicability

The invention, in the various manifestations disclosed herein, has many industrial uses, some of which are the following. The *pilC* DNAs may be used for the design of probes for the detection of *pilC* nucleic acids in samples. The probes derived from the DNAs may be used to detect *pilC* nucleic acids in, for example, chemical synthetic reactions. The polynucleotide probes are also useful in detecting viral nucleic acids in humans, and thus, may serve as a basis for diagnosis of pathogenic microorganisms containing type 4 pilin, for example, gonococcal and/or meningococcal infections in humans.

In addition to the above, the DNAs provided herein provide information and a means for synthesizing polypeptides containing epitopes of PilC. These polypeptides are useful in detecting antibodies to PilC antigens. A series of immunoassays the relevant

neisserial infection, based on recombinant polypeptides containing *pilC* epitopes are described herein, and will find commercial use in diagnosing diseases caused by these microorganisms. In addition, the polypeptides derived from the *pilC* DNAs disclosed herein will have utility as vaccines for treatment of infections caused by meningococci and gonococci.

The polypeptides derived from the *pilC* DNAs, besides the above stated uses, are also useful for raising anti-PilC antibodies. Thus, they may be used in vaccines against the relevant microorganisms. Moreover, the antibodies produced as a result of immunization with the polypeptides containing an immunoreactive PilC epitope are also useful as passive vaccines, or in the detection of the presence of PilC antigens in samples. Thus, they may be used to assay the production of polypeptides derived from PilC in chemical systems. The anti-PilC antibodies may also be used to monitor the efficacy of anti-neisserial agents in screening programs where these agents are tested in tissue culture systems. Another important use for anti-PilC antibodies is in affinity chromatography for the purification of PilC derived polypeptides. The purified PilC polypeptide preparations may be used in vaccines.

For convenience, the anti-PilC antibodies and polypeptides containing regions encoded in *pilC*, whether natural or recombinant, may be packaged into kits.

References cited herein

Badgasarian, M.M., et al., Gene (1983) 26:273-282.

Beard, M.K.M., et al., J. Bacteriol. (1990) 172:2601-2607.

Bergstrom, S., et al., Proc. Natl. Acad. Sci. USA (1986) 83:3890-3894.

Brinton, C.C., et al., in Brooks, G.F., et al. (eds.) Immunobiology of *Neisseria gonorrhoeae* (1978), American Society for Microbiology, Washington, D.C., pp. 155-178.

Edman, P., and Bregg, G., Eur. J. Biochem. (1967) 1:80-91.

Elleman, T.C., and Hoyne, P.A., J. Bacteriol. (1984) 160:1184-1187.

Elleman, T.C., et al., J. Bacteriol. (1986) 168:574-580.

Engstrom, A., et al., EMBO J. (1984) 1:2065-2070.

Fast, R., et al., Gene (1989) 85:227-231.

Fischer, S.H., and Rest, R.F., Infect. Immun. (1988) 56:1574-1579.

Frank-Kamenetskii, M.D., and Vologodskii, A.V., Nature (1984) 304:481-482.

Fürste, J.P., et al., Gene (1986) 48:119-131.

Gill, D.R., et al., Mol. Gen. Genet. (1986) 205:134-145.

Gotschlich, E.C., et al., J. Exp. Med. (1986) 164:868-881.

Haas, R., and Meyer, T., Cell (1986) 44:107-115.

Haas, R., et al., Proc. Natl. Acad. Sci. USA (1987) 84:9079-9083.

Hagblom, P., et al., Nature (1985) 315:156-158.

Hanahan, D., in Glover, D.M. (ed.), DNA Cloning: A Practical Approach (1985), Vol. I., pp. 109-129 (IRL Press, Oxford).

Hanson, M.S., and Brinton, C.C., Nature (1988) 322:265-268.

Hill, S.A., et al., Mol. Microbiol. (1990) 4:1341-1352.

Hultgren, S.J., et al., Proc. Natl. Acad. Sci. USA (1989) 86:4357-4361.

Kellogg, D.S., et al., J. Bacteriol. (1968) 96:596-605.

Laemmli, U.K., Nature (1970), 227:680-685.

Lindberg, F.P., et al., Nature (1987) 328:84-87.

Lindberg et al., J. Bacteriol. (1989) 171:6052-6058.

Marrs, C.F., et al., J. Bacteriol. (1985) 163:132-139.

Mattick, J.S., et al., J. Bacteriol. (1987) 169:33-41.

McKern, N.M., et al., FEBS Lett. (1983) 164:149-153.

Meyer, T., et al., Cell (1982) 30:45-52.

Meyer et al., Proc. Natl. Acad. Sci. USA (1984) 81:6110-6114.

Moch, T., et al., Proc. Natl. Acad. Sci. USA (1987) 84:3462-3466.

Mool, F.R., et al., J. Bacteriol. (1983) 154:41-49.

Murphy, G.L., et al., Cell (1989) 56:589-547.

Norgren, M., et al., Mol. Microbiol. (1987) 1:169-178.

Norlander, L., et al., J. Bacteriol. (1981) 145:788-795.

Norquist, A., et al., FEMS Microbiol. Lett. (1978) 4:71-75.

Orndorff, P.E., and Falkow, S., J. Bacteriol. (1984) 159:736-744.

Panyutin, I.G., et al., Proc. Natl. Acad. Sci. USA (1990) 87:867-870.

Pasloske, B.L., et al., FEBS Lett. (1985) 183:408-412.

Pearce, W.C., and Buchanan, T.M., J. Clin. Invest. (1978) 61:931-943.

Sanger, F., et al., Proc. Natl. Acad. Sci. USA (1977) 74:5463-5467.

Sanger, F., et al., J. Mol. Biol. (1980) 143:161-178.

Segal. E., et al., Cell (1985) 40:293-300.

Seifert, H.S., et al., Proc. Natl. Acad. Sci. USA (1986) 83:735-739.

Seifert, H.S., et al., J. Bacteriol. (1990) 172:40-46.

Shaw, C.E., and Taylor, R.K., Infect. Immun. (1990) 58:3042-3049.

Southern, E.M., J. Mol. Biol. (1975) 98:503-517.

Stern, A., et al., Cell (1986) 47:61-71.

Stibitz, S., et al., Nature (1989) 338:266-269.

Stoker, N.G., et al., in Hames, B.D., and Higgin, S.J. (eds.), Transcription and Translation: A Practical Approach (1984) (IRL Press, Oxford).

Studier et al., Methods in Enzymology (1990) 185:60-89.

Swanson, J., J. Exp. Med. (1973) 137:571-589.

Swanson, J., et al., J. Exp. Med. (1985) 162:729-744.

Swanson, J., et al., Cell (1986) 47:267-276.

Taha, M.K., et al., EMBO J. (1988) 7:4367-4378.

Taylor, T., et al., Vaccine (1988) 6:151-154.

Thompson, R., and Achtman, M., Mol. Gen. Genet. (1978) 165:295-304.

Towbin, H., et al., Proc. Natl. Acad. Sci. USA (1979) 76:4350-4354.

von Heinje, G., Eur. J. Biochem. (1983) 133:17-21.

Watt, P.J., and Ward, M.E., in Beachey, E.H. (ed.), Bacterial Adherence (1980) Series B, Vol. 6, pp. 253-288, Chapman & Hall, London.

Weiser, J.N., et al., Cell (1989) 59:657-665.

Willems, R., et al., EMBO J. (1990) 9:2803-2809.

Yanis-Perron, C., et al., Gene (1985) 33:103-119.

Broach (1981) in Molecular Biology of the Yeast Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.

Broach et al., Meth. Enz. (1983) 101:307.

Catty (1988), ANTIBODIES, Volume 1: A PRACTICAL APPROACH (IRL Press).

Chaney et al. (1986), Cell and Molecular Genetics 12:237.

Chakrabarti et al. (1985), Mol. Cell Biol. 5:3403.

Cohen (1972), Proc. Natl. Acad. Sci. USA 69:2110.

Glennie et al. (1982), Nature 295:712.

Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961.

Grych et al. (1985), Nature 316:74.

Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS.

Hess et al. (1968), J. Adv. Enzyme Reg 7:149.

Hinnen et al. (1978), Proc. Natl. Acad. Sci. 75:1929.

Hitzeman et al. (1980), J. Biol. Chem. 255:2073.

Holland et al. (1978), Biochemistry 17:4900.

Holland (1981), J. Biol. Chem. 256:1385.

Holland and Holland (1980), J. Biol. Chem. 255:2596.

Hunyh, T.V. et al. (1985) in DNA CLONING TECHNIQUES; A PRACTICAL APPROACH (D. Glover, Ed., IRL Press, Oxford, U.K.) pp. 49-78.

Immun. Rev. (1982) 62:185.

Ito et al. (1984), Agric. Biol. Chem. 48:341.

Kennett et al. (1980) MONOCLONAL ANTIBODIES.

Laemmli (1970), Nature 227:680.

Luckow and Summers (1989), Virology 17:31.

Mackett et al. (1984), J. Virol. 49:857.

Maniatis, T., et al. (1982) MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Maniatis et al. (1989), MOLECULAR CLONING; A LABORATORY MANUAL, Second Edition (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London).

Maxam et al. (1980), Methods in Enzymology 65:499.

MacNamara et al. (1984), Science 226:1325.

Messing et al. (1981), Nucleic Acids Res. 9:309.

Messing (1983), Methods in Enzymology 101:20-37.

METHODS IN ENZYMOLOGY (Academic Press).

Moss (1987) in GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (Miller and Calos, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), p. 10.

Neurath et al. (1984), Science 224:392.

Saiki et al. (1986), Nature 324: 163.

Saiki et al. (1988), Science 239:487.

Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463.

Setlow, ed. (1988), GENETIC ENGINEERING. Vol. 10, p. 195-219 (Plenum Publishing Co., N.Y.

Schreier, M., et al. (1980) HYBRIDOMA TECHNIQUES

Scopes (1984), PROTEIN PURIFICATION, PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.).

Tsu and Herzenberg (1980), in SELECTED METHODS IN CELLULAR IMMUNOLOGY (W.H. Freeman and Co.) pp. 373-391.

Vytdehaag et al. (1985), J. Immunol. 134:1225.

Valenzuela, P., et al. (1982), Nature 298:344.

Valenzuela, P., et al. (1984), in HEPATITIS B (Millman, I., et al., ed, Plenum Press) pp. 225-236.

Warner (1984), DNA 3:401.

Wu and Grossman (1987), METHODS IN ENZYMOLOGY Vol. 154, RECOMBINANT DNA, Part E.

Wu (1987), METHODS IN ENZYMOLOGY Vol 155, RECOMBINANT DNA, part F.

Zoller (1982), Nucleic Acids Res. 10:6487.

CLAIMS

1. A recombinant polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

2. The recombinant polynucleotide of claim 1, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

3. The recombinant polynucleotide of claim 2, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

4. The recombinant polynucleotide of claim 2, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

5. A vector comprised of a recombinant polynucleotide, wherein the recombinant polynucleotide is selected from the group consisting of the recombinant polynucleotide of claim 1, the recombinant polynucleotide of claim 2, the recombinant polynucleotide of claim 3, and the recombinant polynucleotide of claim 4.

6. A host cell transformed with the vector of claim 5.

7. A recombinant expression system comprising a polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*, wherein the polynucleotide is operably linked to a control sequence compatible with a desired host.

8. The recombinant expression system of claim 7, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

9. The recombinant expression system of claim 8, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

10. The recombinant expression system of claim 8, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

11. A cell transformed with a recombinant expression system, wherein the expression system is selected from the recombinant expression system of claim 7, the recombinant expression system of claim 8, the recombinant expression system of claim 9, and the recombinant expression system of claim 10.

12. A polypeptide produced by the cell of claim 11.

13. A purified polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

14. A purified polypeptide according to claim 13, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

15. A purified polypeptide according to claim 14, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

16. A purified polypeptide according to claim 14, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

17. A recombinant polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

18. A recombinant polypeptide according to claim 17, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

19. A recombinant polypeptide according to claim 18, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

20. A recombinant polypeptide according to claim 18, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

21. A method of preparing a recombinant polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*, the method comprising:

a. providing a host cell according to claim 11;

b. incubating the cell under conditions which allow expression of the recombinant polypeptide; and

c. isolating the polypeptide.

22. A vaccine composition for the treatment of *Neisseria infection*, comprised of a pharmaceutically acceptable excipient and of an effective amount of a recombinant polypeptide, wherein the polypeptide is

comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

23. The vaccine composition of claim 22, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

24. A polypeptide affixed to a solid substrate, wherein the polypeptide is selected from the group consisting of the polypeptide of claim 12, the polypeptide of claim 13, the polypeptide of claim 14, the polypeptide of claim 15, the polypeptide of claim 16, the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, and the polypeptide of claim 20.

25. An immunoassay for detection of anti-*Neisseria* antibodies comprising:

(a) providing a sample suspected of containing anti-*Neisseria* antibodies;

(b) providing an antigen, wherein the antigen is a polypeptide selected from the group consisting of the polypeptide of claim 12, the polypeptide of claim 13, the polypeptide of claim 14, the polypeptide of claim 15, the polypeptide of claim 16, the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, the polypeptide of claim 20; and

(c) incubating the sample of (a) with the antigen of (b) under conditions which allow the formation of antibody-antigen complexes; and

(d) detecting the presence of anti-*Neisseria* antibody-antigen complexes formed in (c), if any.

26. A composition comprised of a polypeptide, wherein the polypeptide is selected from the group

consisting of the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, and the polypeptide of claim 20.

27. A composition comprised of purified polyclonal anti-PilC antibodies, wherein the PilC is of *Neisseria*.

28. A composition comprised of a monoclonal antibody directed against an immunoreactive epitope encoded in *pilC* of *Neisseria*.

29. An immunoassay for detection of an antigen encoded in *pilC* of *Neisseria* comprising:

(a) providing a sample suspected of containing an antigen encoded in *pilC* of *Neisseria*;

(b) providing a composition comprised of antibodies directed against the antigen encoded in *pilC* of *Neisseria*, wherein the composition is selected from the group of compositions of claim 27 and claim 28;

(c) reacting the sample of (a) with the antibody containing composition of (b) under conditions which allow the formation of anti-PilC antibody-antigen complexes;

(d) detecting anti-PilC antibody-antigen complexes formed in (c), if any.

30. A kit for analyzing samples for the presence of anti-PilC antibodies comprising:

(a) an antigen packaged in a suitable container, wherein the antigen is a polypeptide selected from the group consisting of the polypeptide of claim 12, the polypeptide of claim 13, the polypeptide of claim 14, the polypeptide of claim 15, the polypeptide of claim 16,

the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, and the polypeptide of claim 20;

(b) a buffer used in the performance of the analysis, packaged in a suitable container; and

(c) instructions on the performance of the analysis which uses the antigen of (a) and the buffer of (b).

31. A kit for analyzing samples for the presence of an antigen comprised of an immunoreactive epitope encoded in *pilC* of *Neisseria* comprising:

(a) a composition comprised of antibodies directed against the antigen comprised of an immunoreactive epitope encoded in *pilC* of *Neisseria*, wherein the composition is selected from the group of compositions of claim 27 and claim 28, wherein the composition is packaged in a suitable container;

(b) a buffer used in the performance of the analysis, packaged in a suitable container; and

(c) instructions for performing the analysis.

32. A method for producing antibodies to PilC of *Neisseria* comprising administering to an individual a composition comprised of an isolated immunogenic polypeptide containing a PilC epitope in an amount sufficient to produce an immune response to the PilC epitope.

33. An oligomer capable of hybridizing to a sequence in *pilC* of *Neisseria*, wherein the oligomer is comprised of a *pilC* sequence complementary to at least about 6 contiguous nucleotides of *pilC*.

34. An oligomer according to claim 33, wherein *pilC* is of *Neisseria gonorrheae*.

35. A process for detecting a *pilC* sequence in an analyte strand, wherein the *pilC* sequence comprises a selected target region, the process comprising:

(a) providing a sample comprised of an analyte strand suspected of containing a selected target *pilC* sequence;

(b) providing an oligomer capable of hybridizing to the target *pilC* sequence, wherein the oligomer is comprised of a *pilC* targeting sequence complementary to at least about 6 contiguous nucleotides of *pilC*;

(c) incubating the sample of (a) with the oligomer of (b) under conditions which allow specific hybrid duplexes to form between the targeting sequence and the target sequence; and

(d) detecting hybrids formed between the target sequence, if any, and the oligomer.

36. The process of claim 35 which further comprises:

(a) providing a set of oligomers which are primers for a polymerase chain reaction (PCR) method and which flank the target region; and

(b) amplifying the target region via the PCR method.

37. A kit for detecting a *pilC* sequence in an analyte strand comprising:

(a) the oligomer of claim 33, packaged in a suitable container;

(b) a suitable buffer, packaged in a suitable container; and

(c) instructions for performing the detection.

38. A recombinant polynucleotide comprising a DNA sequence of at least 8 contiguous nucleotides from *pilC*, wherein the *pilC* sequence is selected from the group of sequences shown in Figure 3, Figure 6, and Figure 7.

39. A method of treating an individual for a *Neisseria* infection comprising administering to the individual antibodies produced according to claim 32, wherein the antibodies are administered in an amount effective to prevent the pathology of the infection..

40. An immunoassay for detection of anti-*Neisseria* antibodies comprising:

(a) providing a sample suspected of containing anti-*Neisseria* antibodies;

(b) providing an antigen, wherein the antigen is the polypeptide of claim 24;

(c) incubating the sample of (a) with the antigen of (b) under conditions which allow the formation of antibody-antigen complexes; and

(d) detecting the presence of anti-*Neisseria* antibody-antigen complexes formed in (c), if any.

41. A kit for analyzing samples for the presence of anti-PilC antibodies comprising:

(a) an antigen packaged in a suitable container, wherein the antigen is a polypeptide from claim 24;

(b) a buffer used in the performance of the anaylsis, packaged in a suitable container; and

(c) instructions on the performance of the analysis which uses the antigen of (a) and the buffer of (b).

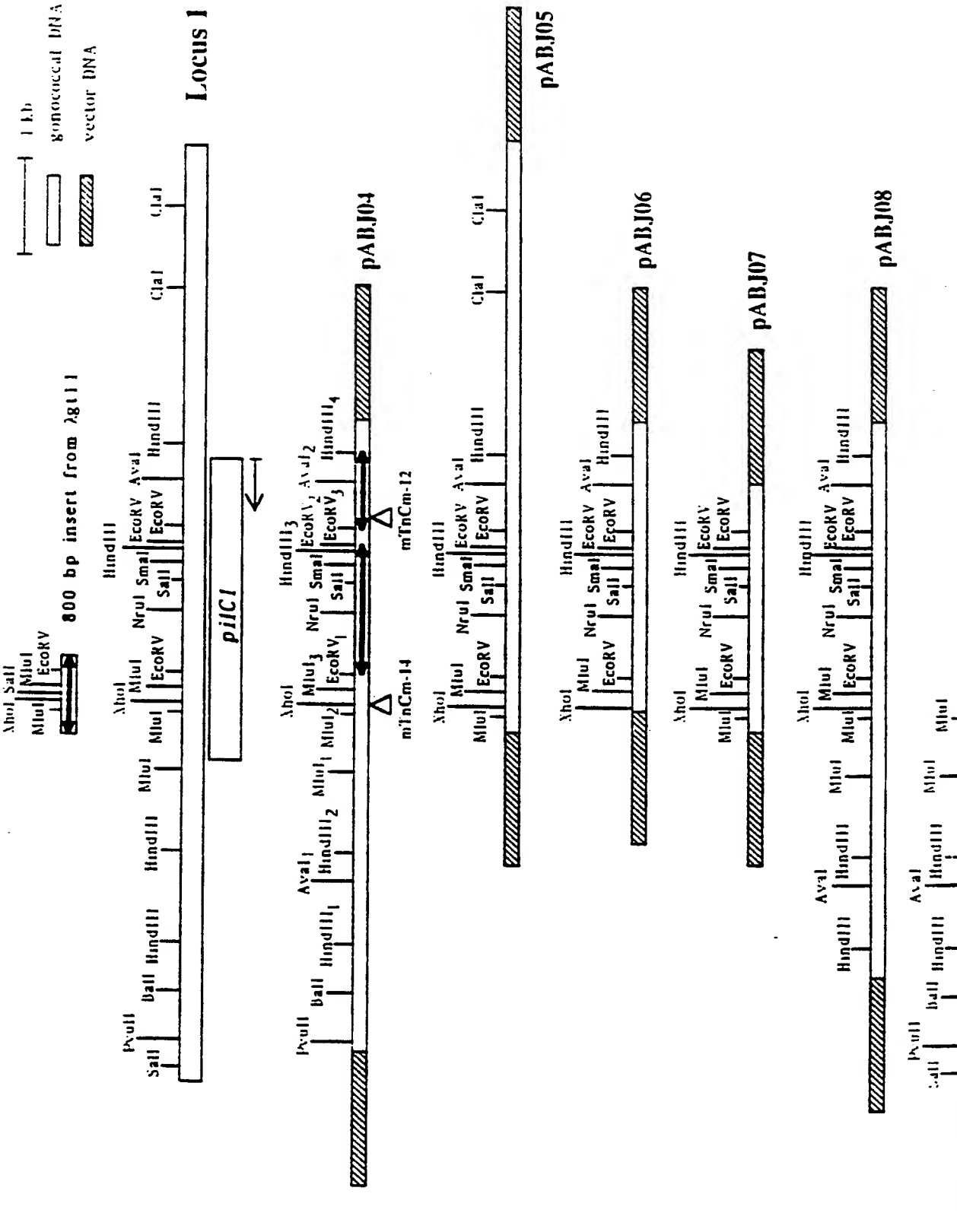

FIGURE 1

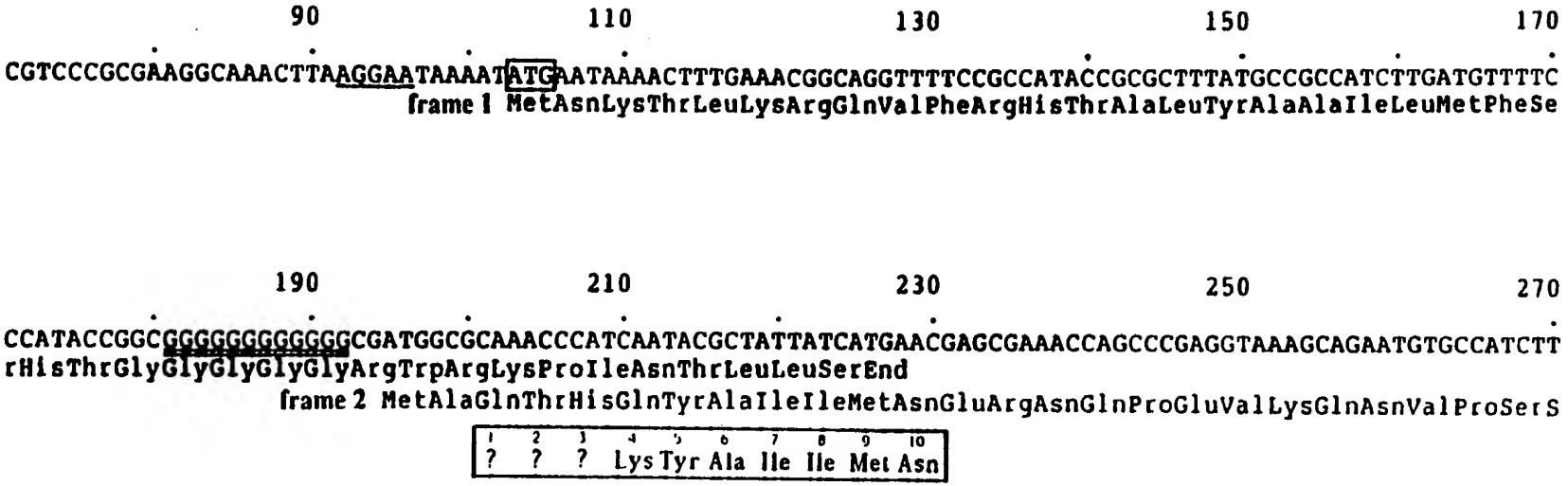

FIGURE 2

```
        10                  30                  50
GATCCGCCCGGTGCTTGGGGCCCTTAGGGAACCGTTCCCTTTGAGCCGGGGCGGGGCAAC
        70                  90                  110
GCGTACCGGTTTTTGTTAATCCGCTATAAAAGGCGGGCTATAGGGTAGGCTTCATCCTG
        130                 150                 170
CCAATCTCACTGAATCCGTCAATTTCCGCAATTCAATTAAATACCGTCAAACCGATGCCG
        190                 210                 230
TCATTCCGCGCAGGCGGGAATCCGGACCGGTCGGGCATCTGCGGCGGTTTGCTAAAAAAC
        250                 270                 290
GCTTTACCGTGATAAGTGCGCAAAGTTAAAATGGGGAGGTAAGCTTTTCAATCAGCAATC
        310                 330                 350
CGGCGGGCGCGGAATCGGGCGGTTTACCGAACCCCGGCGTTCGCGGCGCCCGTCCCGCGA
        370                 390                 410
AGGCAAACTTAAGGAATAAAATATGAATAAAACTTTGAAACGGCAGGTTTTCCGCCATAC
        430                 450                 470
CGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGCGATGGCG
        490                 510                 530
CAAACCCATCAATACGCTATTATCATGAACGAGCGAAACCAGCCCGAGGTAAAGCAGAAT
        550                 570                 590
GTGCCATCTTCAATAAAGGACAAAGACAGGAGGCGCGAATATACTTATTATACGCACAGA
        610                 630                 650
ACAGGAGGAGGCTCTGTCTCATTCAACAATAACGATACCCTTGTTTCCCAACAAAGCGGT
        670                 690                 710
ACTGCCGTTTTTGGCACAGCCACCTACCTGCCGCCCTACGGCAAGGTTTCCGGTTTTGAT
        730                 750                 770
GCCGTCGCTCTGAAAGAGCGCAACAATGCCGTTGATTGGATTCGTACCACCCGCATCGCG
        790                 810                 830
CTGGCAGGCTACTCCTACATCGACGTCATATGCAGAAGCTACACAGGCTGTCCCAAACTT
        850                 870                 890
GTCTATAAAACCCGATTTACCTTCGGTCAACAAGGGTTGAAAAGAAAGGCAGGCAGCAAG
        910                 930                 950
CTGGATATATACGAAGACAAAAGCCGCGAAAATTCGCCCATTTACAAATTGTCGGATTAT
        970                 990                 1010
CCTTGGTTGGGCGTATCTTTCAATTTGGGCAGCGAGAATACCGTCCAAAATAGCAAATTA
        1030                1050                1070
TTCAACAAATTGATATCTTCTTTTAGAGAAGGCAATAATAATCAAACCATCGTCTCTACG
        1090                1110                1130
ACAGAAGGCAACCCTATTTCCCTTGGCGACCGGCAGCGCGAACATACCGCCGTGGCCTAT
        1150                1170                1190
TATCTGAACGCCAAACTGCACCTGCTGGACAAAAAAGGGATTGAAGATATCGCCCAAGGC
        1210                1230                1250
AAAATAGTGGATTTGGGTATCTTGAAACCGCACGTCGAGACGACAGGACGAAGCTTGCTA
        1270                1290                1310
GATTTTTGGGCTAGGTGGGACATTAAAGATACCGGGCAGATTCCGGTCAAGCTCGGCCTG
        1330                1350                1370
CCGCAAGTCAAAGCAGGCCGCTGCACCAACAAACCGAACCCCAATAATAATACCAAAGCC
        1390                1410                1430
CCTTCGCCGGCACTGACCGCCCCCGCGCTGTGGTTCGGACCCGGGCAAGATGGTAAGGCG
        1450                1470                1490
GAGATGTATTCCGCTTCGGTTTCCACCTACCCCGACAGTTCGAGCAGCCGCATCTTCCTC
        1510                1530                1550
CAAGAGCTGAAAACTCAAACCGAACCCGGCAAACCCGGCCGCTATTCCCTCAAATCTTTG
        1570                1590                1610
AATGATGGTGAGATTAAAAGTCGACAGCCGAGTTTCAACGGGCGGCAAACAATCATCCGA
        1630                1650                1670
TTGGATGACGGCGTACATTTGATCAAACTGAATGGAAGCAAGGATGAGGTCGCCGCTTTT
        1690                1710                1730
GTCAATTTAAATGGAAACAACACCGGCAAAAACGACACTTTCGGCATTGTTAAGGAAGCG
        1750                1770                1790
AACGTCAATCTTGACGCCGACGAGTGGAAAAAAGTGCTGCTGCCTTGGACGGTTCGGGGT
        1810                1830                1850
CCCGATAATGACAATAAATTTAAATCAATTAACCAAAAACCAGAAAAATACAGCCAAAGA
        1870                1890                1910
```

FIGURE 3-1

```
TACCGCATCCGCGACAACAACGGCAATCGCGATTTGGGCGACATCGTCAACAGCCCGATT
      1930                1950                1970
GTCGCGGTCGGCGGGTATTTGGCAACCGCCGCGAACGACGGGATGGTGCATATCTTCAAA
      1990                2010                2030
AAAAACGGCGGCAGTGATGAACGCAGCTACAATCTGAAGCTCAGCTACATCCCCGGCACG
      2050                2070                2090
ATGCCGCGCAAGGATATTCAAAGCCAAGAATCCACCCTTGCCAAAGAGCTGCGCGCCTTT
      2110                2130                2150
GCCGAAAAAGGCTATGTGGGCGACCGCTACGGCGTGGACGGCGGCTTTGTCTTGCGCCAA
      2170                2190                2210
GTCGAACTGAGCGGGCAAAAACACGTGTTTATGTTCGGCGCGATGGGTTTTGGCGGCAGG
      2230                2250                2270
GGCGCGTATGCCTTGGATTTAAGCAAAATCAACGGAAATTATCCGGCCGCCGCCCCCCTG
      2290                2310                2330
TTTGATGTCAAAGATGGCGATAATAACGGCAAAAATCGCGTGAAAGTGGAATTAGGCTAC
      2350                2370                2390
ACCGTCGGTACGCCGCAAATCGGCAAAATCCGCAACGGCAAATACGCCGCCTTCCTCGCC
      2410                2430                2450
TCCGGTTATGCGGCTAAAAAAATTGACGACTCAACAAATAAAACCGCGCTGTATGTATAT
      2470                2490                2510
GATTTGAAAGACACCTTAGGTACGCCGATTGCAAAAATCGAAGTGAAGGACGGCAAAGGC
      2530                2550                2570
GGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACGGTCGATATCGCCTAT
      2590                2610                2630
GCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAATTCCGATTCTAGTAAA
      2650                2670                2690
TGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCCGCGCCCGCCGTTTCC
      2710                2730                2750
CGACTGGCAGACAAACGCGTCGTCATCTTCGGTACGGGCAGCGATTTGACCGAAGATGAT
      2770                2790                2810
GTACTGAATACGGGCGAACAATATATTTACGGTATCTTTGACGACGATAAGGGGACGGTT
      2830                2850                2870
AAGGTAACGGTACAAAACGGCACGGCAGGCGGGCTGCTCGAGCAACACCTTACTCAGGAA
      2890                2910                2930
AATAAAACATTATTCCTGAACAAGAGATCCGACGGTTCGGGCAGCAAGGGCTGGGCGGTG
      2950                2970                2990
AAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACCGTGGTATTGCGTACCGCCTTC
      3010                3030                3050
GTAACCATCCGCAAATATAACGACGGCGGCTGCGGCGCGGAAACCGCCATTTTGGGCATC
      3070                3090                3110
AATACCGCCGACGGCGGCGCATTGACTCCGAGAAGCGCGCGCCCGATTGTGCCGGATCAC
      3130                3150                3170
AATTCGGTTGCGCAATATTCCGGCCATAAGACAACCTCCAAAGGCAAATCCATCCCTATA
      3190                3210                3230
GGTTGTATGGACAAAGACGGTAAAACCGTCTGCCCGAACGGATATGTTTACGACAAGCCG
      3250                3270                3290
GTTAATGTGCGTTATCTGGATGAAACGGAAACAGACGGATTTTCAACGACGGCGGACGGC
      3310                3330                3350
GATGCGGGCGGCAGCGGTATAGACCCCGCCGGCAGGCGTCCCGGCAAAAACAACCGCTGC
      3370                3390                3410
TTCTCCAAAAAAGGGGTGCGCACCCTGCTGATGAACGATTTGGACAGCTTGGATATTACC
      3430                3450                3470
GGCCCGATGTGCGGTATCAAACGCTTAAGCTGGCGCGAAGTCTTCTTCTGACCGGCCTGC
      3490                3510                3530
GCGGCCGGTTTTTCCGCAAATGCCGTCCGAAAGGCCTTCGGACGGCATTTTTTTGCGTTT
      3550
TTCGGGAGGGGGGCGGCAAATGAAACG
```

FIGURE 3-2

```
        10                  30                  50
GATCCGCCCGGTGCTTGGGGCCCTTAGGGAACCGTTCCCTTTGAGCCGGGGCGGGGCAAC
        70                  90                 110
GCGTACCGGTTTTTGTTAATCCGCTATAAAAGGCGGGCTATAGGGTAGGCTTCATCCTG
       130                 150                 170
CCAATCTCACTGAATCCGTCAATTTCCGCAATTCAATTAAATACCGTCAAACCGATGCCG
       190                 210                 230
TCATTCCGCGCAGGCGGGAATCCGGACCGGTCGGGCATCTGCGGCGGTTTGCTAAAAAAC
       250                 270                 290
GCTTTACCGTGATAAGTGCGCAAAGTTAAAATGGGGAGGTAAGCTTTTCAATCAGCAATC
       310                 330                 350
CGGCGGGCGCGGAATCGGGCGGTTTACCGAACCCCGGCGTTCGCGGCGCCCGTCCCGCGA
       370                 390                 410
AGGCAAACTTAAGGAATAAAATATGAATAAAACTTTGAAACGGCAGGTTTTCCGCCATAC
                        M   N   K   T   L   K   R   Q   V   F   R   H   T
       430                 450                 470
CGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGGCGATGGC
 A   L   Y   A   A   I   L   M   F   S   H   T   G   G   G   G   G   A   M   A
       490                 510                 530
GCAAACCCATCAATACGCTATTATCATGAACGAGCGAAACCAGCCCGAGGTAAAGCAGAA
 Q   T   H   Q   Y   A   I   I   M   N   E   R   N   Q   P   E   V   K   Q   N
       550                 570                 590
TGTGCCATCTTCAATAAAGGACAAAGACAGGAGGCGCGAATATACTTATTATACGCACAG
 V   P   S   S   I   K   D   K   D   R   R   R   E   Y   T   Y   Y   T   H   R
       610                 630                 650
AACAGGAGGAGGCTCTGTCTCATTCAACAATAACGATACCCTTGTTTCCCAACAAAGCGG
 T   G   G   G   S   V   S   F   N   N   N   D   T   L   V   S   Q   Q   S   G
       670                 690                 710
TACTGCCGTTTTTGGCACAGCCACCTACCTGCCGCCCTACGGCAAGGTTTCCGGTTTTGA
 T   A   V   F   G   T   A   T   Y   L   P   P   Y   G   K   V   S   G   F   D
       730                 750                 770
TGCCGTCGCTCTGAAAGAGCGCAACAATGCCGTTGATTGGATTCGTACCACCCGCATCGC
 A   V   A   L   K   E   R   N   N   A   V   D   W   I   R   T   T   R   I   A
       790                 810                 830
GCTGGCAGGCTACTCCTACATCGACGTCATATGCAGAAGCTACACAGGCTGTCCCAAACT
 L   A   G   Y   S   Y   I   D   V   I   C   R   S   Y   T   G   C   P   K   L
       850                 870                 890
TGTCTATAAAACCCGATTTACCTTCGGTCAACAAGGGTTGAAAAGAAAGGCAGGCAGCAA
 V   Y   K   T   R   F   T   F   G   Q   Q   G   L   K   R   K   A   G   S   K
       910                 930                 950
GCTGGATATATACGAAGACAAAAGCCGCGAAAATTCGCCCATTTACAAATTGTCGGATTA
 L   D   I   Y   E   D   K   S   R   E   N   S   P   I   Y   K   L   S   D   Y
       970                 990                1010
TCCTTGGTTGGGCGTATCTTTCAATTTGGGCAGCGAGAATACCGTCCAAAATAGCAAATT
 P   W   L   G   V   S   F   N   L   G   S   E   N   T   V   Q   N   S   K   L
      1030                1050                1070
ATTCAACAAATTGATATCTTCTTTTAGAGAAGGCAATAATAATCAAACCATCGTCTCTAC
 F   N   K   L   I   S   S   F   R   E   G   N   N   N   Q   T   I   V   S   T
      1090                1110                1130
GACAGAAGGCAACCCTATTTCCCTTGGCGACCGGCAGCGCGAACATACCGCCGTGGCCTA
 T   E   G   N   P   I   S   L   G   D   R   Q   R   E   H   T   A   V   A   Y
      1150                1170                1190
TTATCTGAACGCCAAACTGCACCTGCTGGACAAAAAAGGGATTGAAGATATCGCCCAAGG
 Y   L   N   A   K   L   H   L   L   D   K   K   G   I   E   D   I   A   Q   G
      1210                1230                1250
CAAAATAGTGGATTTGGGTATCTTGAAACCGCACGTCGAGACGACAGGACGAAGCTTGCT
 K   I   V   D   L   G   I   L   K   P   H   V   E   T   T   G   R   S   L   L
```

FIGURE 4-1

```
      1270                1290                1310
AGATTTTTGGGCTAGGTGGGACATTAAAGATACCGGGCAGATTCCGGTCAAGCTCGGCCT
 D  F  W  A  R  W  D  I  K  D  T  G  Q  I  P  V  K  L  G  L
      1330                1350                1370
GCCGCAAGTCAAAGCAGGCCGCTGCACCAACAAACCGAACCCCAATAATAATACCAAAGC
 P  Q  V  K  A  G  R  C  T  N  K  P  N  P  N  N  N  T  K  A
      1390                1410                1430
CCCTTCGCCGGCACTGACCGCCCCCGCGCTGTGGTTCGGACCCGGGCAAGATGGTAAGGC
 P  S  P  A  L  T  A  P  A  L  W  F  G  P  G  Q  D  G  K  A
      1450                1470                1490
GGAGATGTATTCCGCTTCGGTTTCCACCTACCCCGACAGTTCGAGCAGCCGCATCTTCCT
 E  M  Y  S  A  S  V  S  T  Y  P  D  S  S  S  S  R  I  F  L
      1510                1530                1550
CCAAGAGCTGAAAACTCAAACCGAACCCGGCAAACCCGGCCGCTATTCCCTCAAATCTTT
 Q  E  L  K  T  Q  T  E  P  G  K  P  G  R  Y  S  L  K  S  L
      1570                1590                1610
GAATGATGGTGAGATTAAAAGTCGACAGCCGAGTTTCAACGGGCGGCAAACAATCATCCG
 N  D  G  E  I  K  S  R  Q  P  S  F  N  G  R  Q  T  I  I  R
      1630                1650                1670
ATTGGATGACGGCGTACATTTGATCAAACTGAATGGAAGCAAGGATGAGGTCGCCGCTTT
 L  D  D  G  V  H  L  I  K  L  N  G  S  K  D  E  V  A  A  F
      1690                1710                1730
TGTCAATTTAAATGGAAACAACACCGGCAAAAACGACACTTTCGGCATTGTTAAGGAAGC
 V  N  L  N  G  N  N  T  G  K  N  D  T  F  G  I  V  K  E  A
      1750                1770                1790
GAACGTCAATCTTGACGCCGACGAGTGGAAAAAAGTGCTGCTGCCTTGGACGGTTCGGGG
 N  V  N  L  D  A  D  E  W  K  K  V  L  L  P  W  T  V  R  G
      1810                1830                1850
TCCCGATAATGACAATAAATTTAAATCAATTAACCAAAAACCAGAAAAAATACAGCCAAAG
 P  D  N  D  N  K  F  K  S  I  N  Q  K  P  E  K  Y  S  Q  R
      1870                1890                1910
ATACCGCATCCGCGACAACAACGGCAATCGCGATTTGGGCGACATCGTCAACAGCCCGAT
 Y  R  I  R  D  N  N  G  N  R  D  L  G  D  I  V  N  S  P  I
      1930                1950                1970
TGTCGCGGTCGGCGGGTATTTGGCAACCGCCGCGAACGACGGGATGGTGCATATCTTCAA
 V  A  V  G  G  Y  L  A  T  A  A  N  D  G  M  V  H  I  F  K
      1990                2010                2030
AAAAAACGGCGGCAGTGATGAACGCAGCTACAATCTGAAGCTCAGCTACATCCCCGGCAC
 K  N  G  G  S  D  E  R  S  Y  N  L  K  L  S  Y  I  P  G  T
      2050                2070                2090
GATGCCGCGCAAGGATATTCAAAGCCAAGAATCCACCCTTGCCAAAGAGCTGCGCGCCTT
 M  P  R  K  D  I  Q  S  Q  E  S  T  L  A  K  E  L  R  A  F
      2110                2130                2150
TGCCGAAAAAGGCTATGTGGGCGACCGCTACGGCGTGGACGGCGGCTTTGTCTTGCGCCA
 A  E  K  G  Y  V  G  D  R  Y  G  V  D  G  G  F  V  L  R  Q
      2170                2190                2210
AGTCGAACTGAGCGGGCAAAAACACGTGTTTATGTTCGGCGCGATGGGTTTTGGCGGCAG
 V  E  L  S  G  Q  K  H  V  F  M  F  G  A  M  G  F  G  G  R
      2230                2250                2270
GGGCGCGTATGCCTTGGATTTAAGCAAAATCAACGGAAATTATCCGGCCGCCGCCCCCCT
 G  A  Y  A  L  D  L  S  K  I  N  G  N  Y  P  A  A  A  P  L
      2290                2310                2330
GTTTGATGTCAAAGATGGCGATAATAACGGCAAAAATCGCGTGAAAGTGGAATTAGGCTA
 F  D  V  K  D  G  D  N  N  G  K  N  R  V  K  V  E  L  G  Y
      2350                2370                2390
CACCGTCGGTACGCCGCAAATCGGCAAAATCCGCAACGGCAAATACGCCGCCTTCCTCGC
 T  V  G  T  P  Q  I  G  K  I  R  N  G  K  Y  A  A  F  L  A
      2410                2430                2450
CTCCGGTTATGCGGCTAAAAAAATTGACGACTCAACAAATAAAACCGCGCTGTATGTATA
 S  G  Y  A  A  K  K  I  D  D  S  T  N  K  T  A  L  Y  V  Y
      2470                2490                2510
TGATTTGAAAGACACCTTAGGTACGCCGATTGCAAAAATCGAAGTGAAGGACGGCAAAGG
 D  L  K  D  T  L  G  T  P  I  A  K  I  E  V  K  D  G  K  G
```

FIGURE 4-2

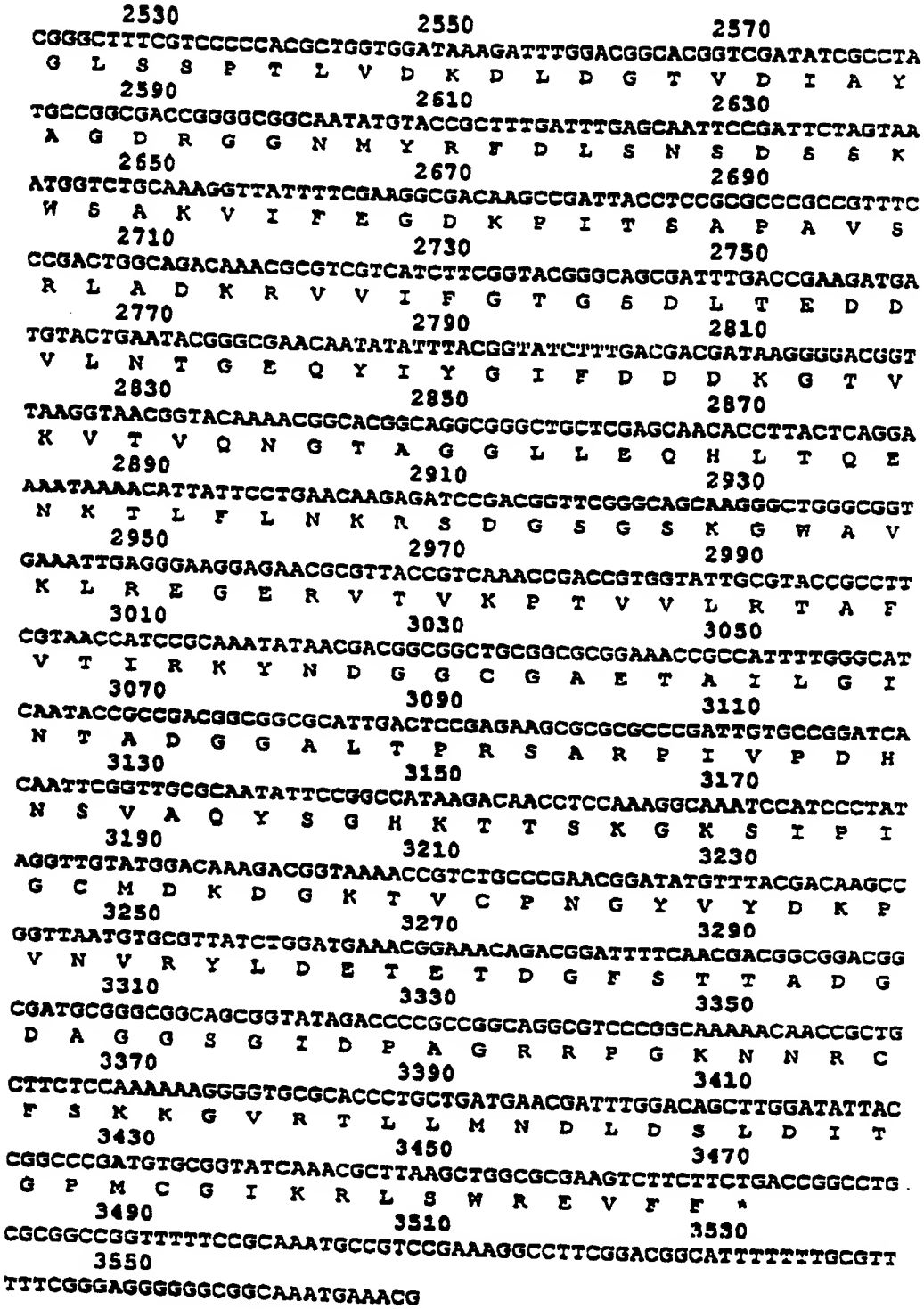

FIGURE 4-3

```
        10                  30                  50
GATCCGCCCGGTGCTTGGGGCCCTTAGGGAACCGTTCCCTTTGAGCCGGGGCGGGGCAAC
        70                  90                  110
GCGTACCGGTTTTTGTTAATCCGCTATAAAAGGCGGGCTATAGGGTAGGCTTCATCCTG
        130                 150                 170
CCAATCTCACTGAATCCGTCAATTTCCGCAATTCAATTAAATACCGTCAAACCGATGCCG
        190                 210                 230
TCATTCCGCGCAGGCGGGAATCCGGACCGGTCGGGCATCTGCGGCGGTTTGCTAAAAAAC
        250                 270                 290
GCTTTACCGTGATAAGTGCGCAAAGTTAAAATGGGGAGGTAAGCTTTTCAATCAGCAATC
        310                 330     350
CGGCGGGCGCGGAATCGGGCGGTTTACCGAACCCCGGCGTTCGCGGCGCCCGTCCCGCGA
        370                 390                 410
AGGCAAACTTAAGGAATAAAATATGAATAAAACTTTGAAACGGCAGGTTTTCCGCCATAC
            frame 1   M  N  K  T  L  K  R  Q  V  F  R  H  T
        430                 450                 470
CGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGCGATGGCG
 A  L  Y  A  A  I  L  M  F  S  H  T  G  G  G  G  G  R  W  R
                                                frame 2  M  A
        490                 510                 530
CAAACCCATCAATACGCTATTATCATGAACGAGCGAAACCAGCCCGAGGTAAAGCAGAAT
 K  P  I  N  T  L  L  S  *
Q  T  H  Q  Y  A  I  I  M  N  E  R  N  Q  P  E  V  K  Q  N
        550                 570                 590
GTGCCATCTTCAATAAAGGACAAAGACAGGAGGCGCGAATATACTTATTATACGCACAGA
V  P  S  S  I  K  D  K  D  R  R  R  E  Y  T  Y  Y  T  H  R
        610                 630                 650
ACAGGAGGAGGCTCTGTCTCATTCAACAATAACGATACCCTTGTTTCCCAACAAAGCGGT
T  G  G  G  S  V  S  F  N  N  N  D  T  L  V  S  Q  Q  S  G
        670                 690                 710
ACTGCCGTTTTTGGCACAGCCACCTACCTGCCGCCCTACGGCAAGGTTTCCGGTTTTGAT
T  A  V  F  G  T  A  T  Y  L  P  P  Y  G  K  V  S  G  F  D
        730                 750                 770
GCCGTCGCTCTGAAAGAGCGCAACAATGCCGTTGATTGGATTCGTACCACCCGCATCGCG
A  V  A  L  K  E  R  N  N  A  V  D  W  I  R  T  T  R  I  A
        790                 810                 830
CTGGCAGGCTACTCCTACATCGACGTCATATGCAGAAGCTACACAGGCTGTCCCAAACTT
L  A  G  Y  S  Y  I  D  V  I  C  R  S  Y  T  G  C  P  K  L
        850                 870                 890
GTCTATAAAACCCGATTTACCTTCGGTCAACAAGGGTTGAAAAGAAAGGCAGGCAGCAAG
V  Y  K  T  R  F  T  F  G  Q  Q  G  L  K  R  K  A  G  S  K
        910                 930                 950
CTGGATATATACGAAGACAAAAGCCGCGAAAATTCGCCCATTTACAAATTGTCGGATTAT
L  D  I  Y  E  D  K  S  R  E  N  S  P  I  Y  K  L  S  D  Y
        970                 990                 1010
CCTTGGTTGGGCGTATCTTTCAATTTGGGCAGCGAGAATACCGTCCAAAATAGCAAATTA
P  W  L  G  V  S  F  N  L  G  S  E  N  T  V  Q  N  S  K  L
        1030                1050                1070
TTCAACAAATTGATATCTTCTTTTAGAGAAGGCAATAATAATCAAACCATCGTCTCTACG
F  N  K  L  I  S  S  F  R  E  G  N  N  N  Q  T  I  V  S  T
        1090                1110                1130
ACAGAAGGCAACCCTATTTCCCTTGGCGACCGGCAGCGCGAACATACCGCCGTGGCCTAT
T  E  G  N  P  I  S  L  G  D  R  Q  R  E  H  T  A  V  A  Y
        1150                1170                1190
TATCTGAACGCCAAACTGCACCTGCTGGACAAAAAAGGGATTGAAGATATCGCCCAAGGC
Y  L  N  A  K  L  H  L  L  D  K  K  G  I  E  D  I  A  Q  G
        1210                1230                1250
AAAATAGTGGATTTGGGTATCTTGAAACCGCACGTCGAGACGACAGGACGAAGCTTGCTA
K  I  V  D  L  G  I  L  K  P  H  V  E  T  T  G  R  S  L  L
        1270                1290                1310
GATTTTTGGGCTAGGTGGGACATTAAAGATACCGGGCAGATTCCGGTCAAGCTCGGCCTG
D  F  W  A  R  W  D  I  K  D  T  G  Q  I  P  V  K  L  G  L
```

FIGURE 5-1

```
      1330                 1350                 1370
CCGCAAGTCAAAGCAGGCCGCTGCACCAACAAACCGAACCCCAATAATAATACCAAAGCC
P  Q  V  K  A  G  R  C  T  N  K  P  N  P  N  N  N  T  K  A
      1390                 1410                 1430
CCTTCGCCGGCACTGACCGCCCCCGCGCTGTGGTTCGGACCCGGGCAAGATGGTAAGGCG
P  S  P  A  L  T  A  P  A  L  W  F  G  P  G  Q  D  G  K  A
      1450                 1470                 1490
GAGATGTATTCCGCTTCGGTTTCCACCTACCCCGACAGTTCGAGCAGCCGCATCTTCCTC
E  M  Y  S  A  S  V  S  T  Y  P  D  S  S  S  S  R  I  F  L
      1510                 1530                 1550
CAAGAGCTGAAAACTCAAACCGAACCCGGCAAACCCGGCCGCTATTCCCTCAAATCTTTG
Q  E  L  K  T  Q  T  E  P  G  K  P  G  R  Y  S  L  K  S  L
      1570                 1590                 1610
AATGATGGTGAGATTAAAAGTCGACAGCCGAGTTTCAACGGGCGGCAAACAATCATCCGA
N  D  G  E  I  K  S  R  Q  P  S  F  N  G  R  Q  T  I  I  R
      1630                 1650                 1670
TTGGATGACGGCGTACATTTGATCAAACTGAATGGAAGCAAGGATGAGGTCGCCGCTTTT
L  D  D  G  V  H  L  I  K  L  N  G  S  K  D  E  V  A  A  F
      1690                 1710                 1730
GTCAATTTAAATGGAAACAACACCGGCAAAAACGACACTTTCGGCATTGTTAAGGAAGCG
V  N  L  N  G  N  N  T  G  K  N  D  T  F  G  I  V  K  E  A
      1750                 1770                 1790
AACGTCAATCTTGACGCCGACGAGTGGAAAAAAGTGCTGCTGCCTTGGACGGTTCGGGGT
N  V  N  L  D  A  D  E  W  K  K  V  L  L  P  W  T  V  R  G
      1810                 1830                 1850
CCCGATAATGACAATAAATTTAAATCAATTAACCAAAAACCAGAAAAATACAGCCAAAGA
P  D  N  D  N  K  F  K  S  I  N  Q  K  P  E  K  Y  S  Q  R
      1870                 1890                 1910
TACCGCATCCGCGACAACAACGGCAATCGCGATTTGGGCGACATCGTCAACAGCCCGATT
Y  R  I  R  D  N  N  G  N  R  D  L  G  D  I  V  N  S  P  I
      1930                 1950                 1970
GTCGCGGTCGGCGGGTATTTGGCAACCGCCGCGAACGACGGGATGGTGCATATCTTCAAA
V  A  V  G  G  Y  L  A  T  A  A  N  D  G  M  V  H  I  F  K
      1990                 2010                 2030
AAAAACGGCGGCAGTGATGAACGCAGCTACAATCTGAAGCTCAGCTACATCCCCGGCACG
K  N  G  G  S  D  E  R  S  Y  N  L  K  L  S  Y  I  P  G  T
      2050                 2070                 2090
ATGCCGCGCAAGGATATTCAAAGCCAAGAATCCACCCTTGCCAAAGAGCTGCGCGCCTTT
M  P  R  K  D  I  Q  S  Q  E  S  T  L  A  K  E  L  R  A  F
      2110                 2130                 2150
GCCGAAAAAGGCTATGTGGGCGACCGCTACGGCGTGGACGGCGGCTTTGTCTTGCGCCAA
A  E  K  G  Y  V  G  D  R  Y  G  V  D  G  G  F  V  L  R  Q
      2170                 2190                 2210
GTCGAACTGAGCGGGCAAAAACACGTGTTTATGTTCGGCGCGATGGGTTTTGGCGGCAGG
V  E  L  S  G  Q  K  H  V  F  M  F  G  A  M  G  F  G  G  R
      2230                 2250                 2270
GGCGCGTATGCCTTGGATTTAAGCAAAATCAACGGAAATTATCCGGCCGCCGCCCCCCTG
G  A  Y  A  L  D  L  S  K  I  N  G  N  Y  P  A  A  A  P  L
      2290                 2310                 2330
TTTGATGTCAAAGATGGCGATAATAACGGCAAAAATCGCGTGAAAGTGGAATTAGGCTAC
F  D  V  K  D  G  D  N  N  G  K  N  R  V  K  V  E  L  G  Y
      2350                 2370                 2390
ACCGTCGGTACGCCGCAAATCGGCAAAATCCGCAACGGCAAATACGCCGCCTTCCTCGCC
T  V  G  T  P  Q  I  G  K  I  R  N  G  K  Y  A  A  F  L  A
      2410                 2430                 2450
TCCGGTTATGCGGCTAAAAAAATTGACGACTCAACAAATAAAACCGCGCTGTATGTATAT
S  G  Y  A  A  K  K  I  D  D  S  T  N  K  T  A  L  Y  V  Y
      2470                 2490                 2510
GATTTGAAAGACACCTTAGGTACGCCGATTGCAAAAATCGAAGTGAAGGACGGCAAAGGC
D  L  K  D  T  L  G  T  P  I  A  K  I  E  V  K  D  G  K  G
      2530                 2550                 2570
GGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACGGTCGATATCGCCTAT
G  L  S  S  P  T  L  V  D  K  D  L  D  G  T  V  D  I  A  Y
```

FIGURE 5-2

```
      2590                2610                2630
GCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAATTCCGATTCTAGTAAA
A  G  D  R  G  G  N  M  Y  R  F  D  L  S  N  S  D  S  S  K
      2650                2670                2690
TGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCCGCGCCCGCCGTTTCC
W  S  A  K  V  I  F  E  G  D  K  P  I  T  S  A  P  A  V  S
      2710                2730                2750
CGACTGGCAGACAAACGCGTCGTCATCTTCGGTACGGGCAGCGATTTGACCGAAGATGAT
R  L  A  D  K  R  V  V  I  F  G  T  G  S  D  L  T  E  D  D
      2770                2790                2810
GTACTGAATACGGGCGAACAATATATTTACGGTATCTTTGACGACGATAAGGGGACGGTT
V  L  N  T  G  E  Q  Y  I  Y  G  I  F  D  D  D  K  G  T  V
      2830                2850                2870
AAGGTAACGGTACAAAACGGCACGGCAGGCGGGCTGCTCGAGCAACACCTTACTCAGGAA
K  V  T  V  Q  N  G  T  A  G  G  L  L  E  Q  H  L  T  Q  E
      2890                2910                2930
AATAAAACATTATTCCTGAACAAGAGATCCGACGGTTCGGGCAGCAAGGGCTGGGCGGTG
N  K  T  L  F  L  N  K  R  S  D  G  S  G  S  K  G  W  A  V
      2950                2970                2990
AAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACCGTGGTATTGCGTACCGCCTTC
K  L  R  E  G  E  R  V  T  V  K  P  T  V  V  L  R  T  A  F
      3010                3030                3050
GTAACCATCCGCAAATATAACGACGGCGGCTGCGGCGCGGAAACCGCCATTTTGGGCATC
V  T  I  R  K  Y  N  D  G  G  C  G  A  E  T  A  I  L  G  I
      3070                3090                3110
AATAGCGCCGGACGCGCGGCGCATTGACTCCGAGAAGCGCGCGCCCGATTGTGCCGGATCAC
N  T  A  D  G  G  A  L  T  P  R  S  A  R  P  I  V  P  D  H
      3130                3150                3170
AATTCGGTTGCGCAATATTCCGGCCATAAGACAACCTCCAAAGGCAAATCCATCCCTATA
N  S  V  A  Q  Y  S  G  H  K  T  T  S  K  G  K  S  I  P  I
      3190                3210                3230
GGTTGTATGGACAAAGACGGTAAAACCGTCTGCCCGAACGGATATGTTTACGACAAGCCG
G  C  M  D  K  D  G  K  T  V  C  P  N  G  Y  V  Y  D  K  P
      3250                3270                3290
GTTAATGTGCGTTATCTGGATGAAACGGAAACAGACGGATTTTCAACGACGGCGGACGGC
V  N  V  R  Y  L  D  E  T  E  T  D  G  F  S  T  T  A  D  G
      3310                3330                3350
GATGCGGGCGGCAGCGGTATAGACCCCGCCGGCAGGCGTCCCGGCAAAAACAACCGCTGC
D  A  G  G  S  G  I  D  P  A  G  R  R  P  G  K  N  N  R  C
      3370                3390                3410
TTCTCCAAAAAAGGGGTGCGCACCCTGCTGATGAACGATTTGGACAGCTTGGATATTACC
F  S  K  K  G  V  R  T  L  L  M  N  D  L  D  S  L  D  I  T
      3430                3450                3470
GGCCCGATGTGCGGTATCAAACGCTTAAGCTGGCGCGAAGTCTTCTTCTGACCGGCCTGC
G  P  M  C  G  I  K  R  L  S  W  R  E  V  F  F  *
      3490                3510                3530
GCGGCCGGTTTTTCCGCAAATGCCGTCCGAAAGGCCTTCGGACGGCATTTTTTTGCGTTT
      3550
TTCGGGAGGGGGGCGGCAAATGAAACG
```

FIGURE 5-3

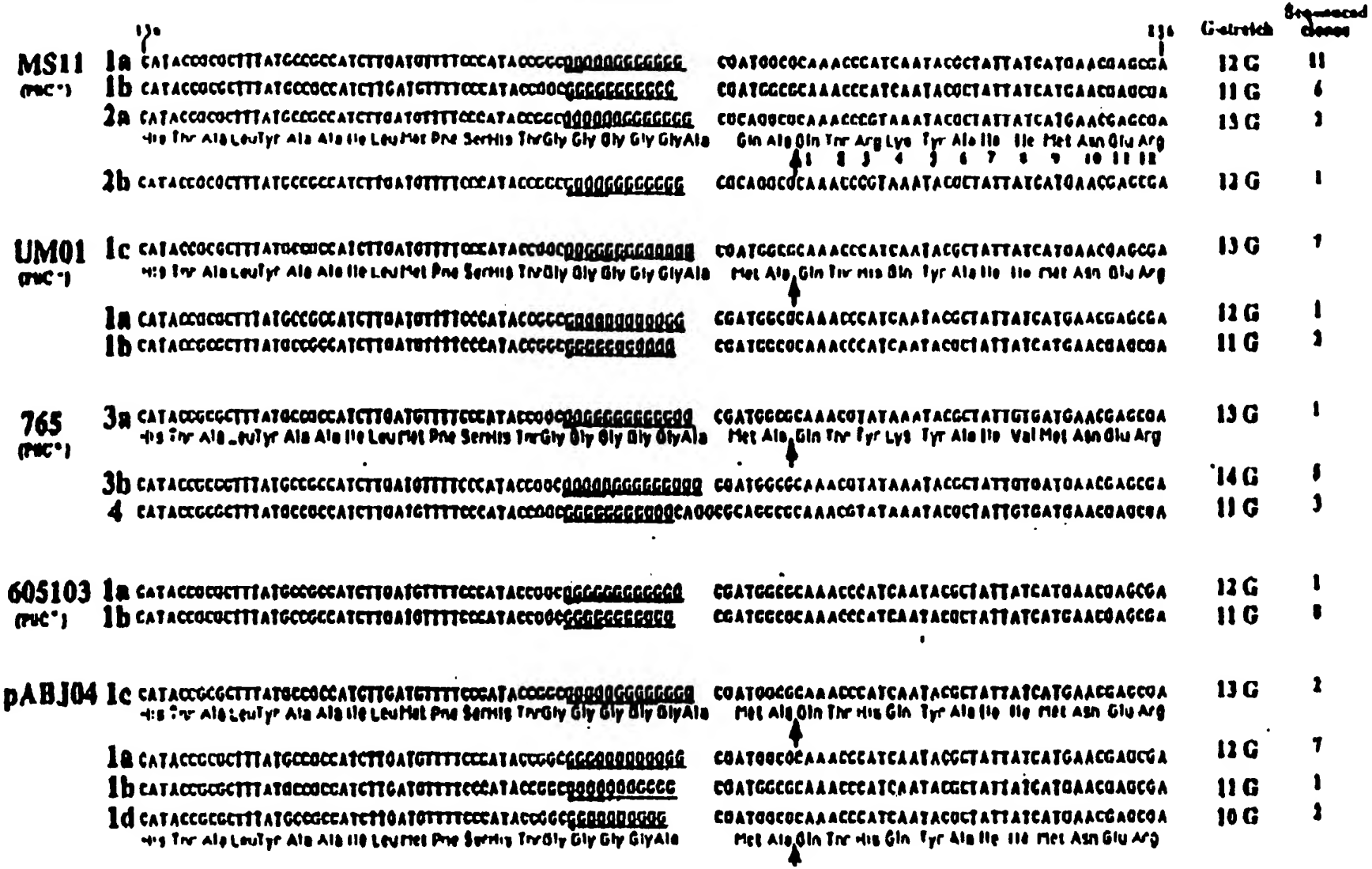
MS11 (pUC*)
1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
2a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGCAGGCGCAAACCCGTAAATACGCTATTATCATGAACGAGCGA
His Thr Ala LeuTyr Ala Ala Ile LeuMet Phe SerHis ThrGly Gly Gly Gly GlyAla Gln Ala Gln Thr Arg Lys Tyr Ala Ile Ile Met Asn Glu Arg
1 2 3 4 5 6 7 8 9 10 11 12
2b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGG CGCAGGCGCAAACCCGTAAATACGCTATTATCATGAACGAGCGA
UM01 (pUC*)
1c CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
His Thr Ala LeuTyr Ala Ala Ile LeuMet Phe SerHis ThrGly Gly Gly Gly GlyAla Met Ala Gln Thr His Gln Tyr Ala Ile Ile Met Asn Glu Arg
1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
765 (pUC*)
3a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGATGGCGCAAACGTATAAATACGCTATTGTGATGAACGAGCGA
His Thr Ala LeuTyr Ala Ala Ile LeuMet Phe SerHis ThrGly Gly Gly Gly GlyAla Met Ala Gln Thr Tyr Lys Tyr Ala Ile Val Met Asn Glu Arg
3b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGATGGCGCAAACGTATAAATACGCTATTGTGATGAACGAGCGA
4 CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGCAGCGCAGCGCCAAACGTATAAATACGCTATTGTGATGAACGAGCGA
605103 (pUC*)
1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
pABJ04 1c CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
His Thr Ala LeuTyr Ala Ala Ile LeuMet Phe SerHis ThrGly Gly Gly Gly GlyAla Met Ala Gln Thr His Gln Tyr Ala Ile Ile Met Asn Glu Arg
1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
1d CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGCCGGGGGGGGGG CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA
His Thr Ala LeuTyr Ala Ala Ile LeuMet Phe SerHis ThrGly Gly Gly GlyAla Met Ala Gln Thr His Gln Tyr Ala Ile Ile Met Asn Glu Arg
G-stretch
12 G
11 G
13 G
12 G
11 G
13 G
14 G
11 G
12 G
11 G
13 G
12 G
11 G
10 G
Sequenced clones
11
6
3
1
7
1
1
1
1
1
2
7
1
3

FIGURE 6

1 CCGCTGTATGTGTATGATTTGGAAAACACCAGTGGTAGTCTGATTAAAAAAATCGAAGCA

CCCGGCGGCAAAGGCGGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACG

GTCGATATCGCCTATGCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAAT

TCCGATTCTAGTAAATGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCC

GCGCCCGCCGTTTCCCGACTGGCAGACAAACGCGTGGTTATCTTCGGCACGGGCAGCGAT

TTGAGTGAACAGGATGTACTGGATACGGACAAACAATATATTTACGGTATCTTTGACGAC

GATAAGTCGACGGTTAATGTAAAGGTAACAAACGGCACGGGAGGCGGGCTGCTCGAGCAA

GTGCTTAAAGAGGAAAGTAAAACCTTATTCCTGAGCAATAATAAGGCATCCGGCGGATCG

GCCGATAAAGGGTGGGTAGTGAAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACC

GTGGTATTGCGTACCGCCTTTGTCACCATCCGCAAATATACGGATACGGACAAATGTGGC

GCGCAAACCGCCATTTTGGGCATCAATACCGCCGACGGCGGCGCATTGACTCCGAGAAGC

GCGCGCCCGATTGTGCCGGATCACAATTCGGTTGCGCAATATTCCGGCCATCAGAAAATG

AACGGCAAGTCCATCCCGG 739

FIGURE 7

```
   P  L  Y  V  Y  D  L  E  N  T  S  G  S  L  I  K  K  I  E  A
1  CCGCTGTATGTGTATGATTTGGAAAACACCAGTGGTAGTCTGATTAAAAAAATCGAAGCA

   P  G  G  K  G  G  L  S  S  P  T  L  V  D  K  D  L  D  G  T
   CCCGGCGGCAAAGGCGGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACG

   V  D  I  A  Y  A  G  D  R  G  G  N  M  Y  R  F  D  L  S  N
   GTCGATATCGCCTATGCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAAT

   S  D  S  S  K  W  S  A  K  V  I  F  E  G  D  K  P  I  T  S
   TCCGATTCTAGTAAATGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCC

   A  P  A  V  S  R  L  A  D  K  R  V  V  I  F  G  T  G  S  D
   GCGCCCGCCGTTTCCCGACTGGCAGACAAACGCGTGGTTATCTTCGGCACGGGCAGCGAT

   L  S  E  Q  D  V  L  D  T  D  K  Q  Y  I  Y  G  I  F  D  D
   TTGAGTGAACAGGATGTACTGGATACGGACAAACAATATATTTACGGTATCTTTGACGAC

   D  K  S  T  V  N  V  K  V  T  N  G  T  G  G  G  L  L  E  Q
   GATAAGTCGACGGTTAATGTAAAGGTAACAAACGGCACGGGAGGCGGGCTGCTCGAGCAA

   V  L  K  E  E  S  K  T  L  F  L  S  N  N  K  A  S  G  G  S
   GTGCTTAAAGAGGAAAGTAAAACCTTATTCCTGAGCAATAATAAGGCATCCGGCGGATCG

   A  D  K  G  W  V  V  K  L  R  E  G  E  R  V  T  V  K  P  T
   GCCGATAAAGGGTGGGTAGTGAAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACC

   V  V  L  R  T  A  F  V  T  I  R  K  Y  T  D  T  D  K  C  G
   GTGGTATTGCGTACCGCCTTTGTCACCATCCGCAAATATACGGATACGGACAAATGTGGC

   A  Q  T  A  I  L  G  I  N  T  A  D  G  G  A  L  T  P  R  S
   GCGCAAACCGCCATTTTGGGCATCAATACCGCCGACGGCGGCGCATTGACTCCGAGAAGC

   A  R  P  I  V  P  D  H  N  S  V  A  Q  Y  S  G  H  Q  K  M
   GCGCGCCCGATTGTGCCGGATCACAATTCGGTTGCGCAATATTCCGGCCATCAGAAAATG.

   N  G  K  S  I  P
   AACGGCAAGTCCATCCCGG  739
```

FIGURE 8

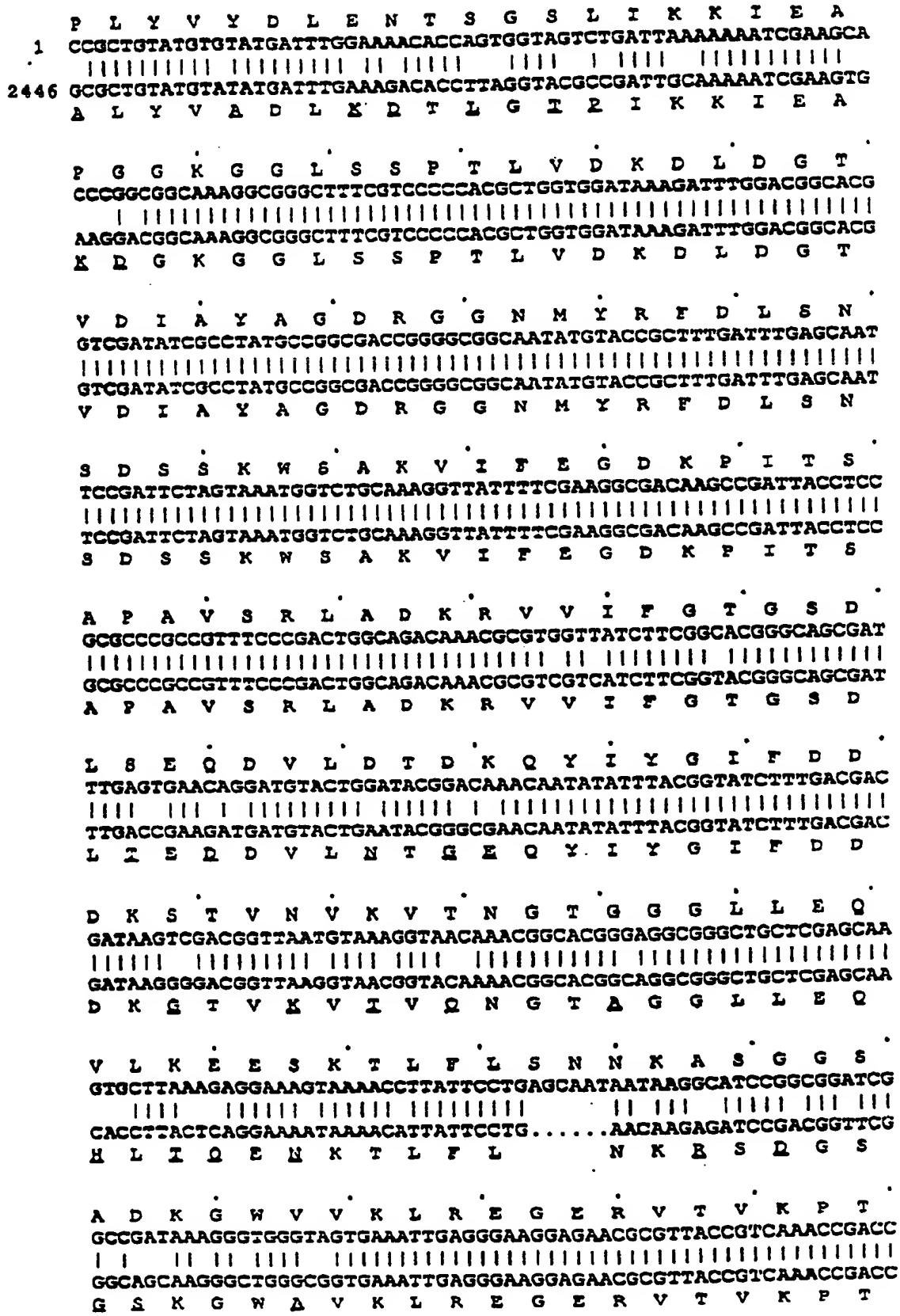

FIGURE 9-1

```
V  V  L  R  T  A  F  V  T  I  R  K  Y  T  D  T  D  K  C  G
GTGGTATTGCGTACCGCCTTTGTCACCATCCGCAAATATACGGATACGGACAAATGTGGC
||||||||||||||||||| || |||||||||||||||||  ||    | |   || |||
GTGGTATTGCGTACCGCCTTCGTAACCATCCGCAAATATAACGA...CGGCGGCTGCGGC
V  V  L  R  T  A  F  V  T  I  R  K  Y  N  D     G  G  C  G

A  Q  T  A  I  L  G  I  N  T  A  D  G  G  A  L  T  P  R  S
GCGCAAACCGCCATTTTGGGCATCAATACCGCCGACGGCGGCGCATTGACTCCGAGAAGC
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GCGGAAACCGCCATTTTGGGCATCAATACCGCCGACGGCGGCGCATTGACTCCGAGAAGC
A  E  T  A  I  L  G  I  N  T  A  D  G  G  A  L  T  P  R  S

A  R  P  I  V  P  D  H  N  S  V  A  Q  Y  S  G  H  Q  K  M
GCGCGCCCGATTGTGCCGGATCACAATTCGGTTGCGCAATATTCCGGCCATCAGAAAA.
|||||||||||||||||||||||||||||||||||||||||||||||||| ||| ||
GCGCGCCCGATTGTGCCGGATCACAATTCGGTTGCGCAATATTCCGGCCATAAGACAAC
A  R  P  I  V  P  D  H  N  S  V  A  Q  Y  S  G  H  K  T  T

   N  G  K  S  I  P
..TGAACGGCAAGTCCATCCCGG  739
  ||||  |||||  ||||||||
CTCCAAAGGCAAATCCATCCCAT  3176
 S  K  G  K  S  I  P
```

FIGURE 9-2

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00863

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C07H 21/00; G01N 33/53
US CL : 536/27; 435/7.3

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	536/27-29; 435/6,7.1,7.2,7.3,7.36,69.1, 69.3,91,243,871,961; 436/501,518,547.811; 530/300,324,333,350,388 4,339.5,403,825: 935/9, 12, 15, 66, 72, 78, 81

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

CAS ONLINE, MEDLINE, APS, BIOSIS

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
X/Y	US,A, 4,443,431 (BUCHANAN ET AL.) 17 APRIL 1984, SEE ESPECIALLY COLUMN 3, LINES 1-7.	1 - 4 , 7 - 10,33,34,38/5, 6 , 1 1 , 1 2 - 24,26,30,37,41
A	JOURNAL OF BACTERIOLOGY, VOLUME 130, NUMBER 1, ISSUED APRIL 1977, SWANEY ET AL., "GENETIC COMPLEMENTATION ANALYSIS OF ESCHERICHIA COLI TYPE 1 SOMATIC PILUS MUTANTS", PAGES 506-511, SEE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41
X/Y	US,A, 4,584,195 (SCHOOLNIK ET AL.) 22 APRIL 1986, SEE ESPECIALLY THE ABSTRACT AND CLAIMS 1-14.	1 - 4 , 7 - 10,22,23/5,6,1 1,12,24,30,33, 34,37,38,41
Y	JOURNAL OF BACTERIOLOGY, VOLUME 172, NUMBER 6, ISSUED JUNE 1990, NUNN ET AL., "PRODUCTS OF THREE ACCESSORY GENES, PILB, PILC, PILD, ARE REQUIRED FOR BIOGENESIS OF PSEUDOMONAS AERUGINOSA PILI", PAGES 2911-2919, SEE ESPECIALLY THE ABSTRACT.	1 - 24,26,30,33,34 ,37,38, 41

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
27 MAY 1992	02 JUN 1992
International Searching Authority[1]	Signature of Authorized Officer[20]
ISA/US	Edward Fruse for. ARDIN H. MARSCHEL

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	JOURNAL OF BACTERIOLOGY, VOLUME 170, NUMBER 4, ISSUED APRIL 1988, PERRY ET AL., "NEISSERIA MENINGITIDIS C114 CONTAINS SILENT, TRUNCATED PILIN GENES THAT ARE HOMOLOGOUS TO NEISSERIA GONORRHOEAE PIL SEQUENCES", PAGES 1691-1697, SEE ESPECIALLY PAGE 1691, SECOND COLUMN, LINES 3-8.	1 24,26,30,33,34 ,37,38,41
X/Y	THE JOURNAL OF BIOLOGICAL CHEMISTRY, VOLUME 254, NUMBER 9, ISSUED 10 MAY 1979, KELLEY ET AL., "A RAPID PROCEDURE FOR ISOLATION OF LARGE QUANTITIES OF ESCHERICHIA COLI DNA POLYMERASE I UTILIZING A LAMBDA-POL A TRANSDUCING PHAGE", PAGES 3206-3210, SEE ESPECIALLY TABLE II ON PAGE 3208.	12,24/30,41

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE[1]

1. ☐ Claim numbers _, because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers _, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers _, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING[2]

This International Searching Authority found multiple inventions in this international application as follows:

Please See Attached Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

1-24,26,30,33,34,37,38,41 (Telephone Practice) (Telephone Practice) (Telephone Practice)

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No.[18]
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (USA), VOLUME 84, ISSUED DECEMBER 1987, HAAS ET AL., "RELEASE OF SOLUBLE PILIN ANTIGEN COUPLED WITH GENE CONVERSION IN NEISSERIA GONORRHOEAE", PAGES 9079-9083, SEE THE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41
X/Y	NEW ENGLAND BIOLABS CATALOG, ISSUED 1986, (NEW ENGLAND BIOLABS, BEVERLY, MASSACHUSETTS, 1986), PAGE 60, SEE ESPECIALLY LINKER # 1096 COMPARED TO THE INSTANT APPLICATION FIGURE 3 AT BASES 3196-3202.	33,34/37
X/Y	SIGMA CHEMICAL COMPANY CATALOG, ISSUED 1990, (SIGMA CHEMICAL COMPANY, ST. LOUIS, MISSOURI, 1990), PAGES 859-860, SEE ESPECIALLY POLY[C]-[dG]12-18 ON PAGE 859 AND POLYDEOXYGUANYLIC ACID ON PAGE 860 COMPARED TO THE INSTANT APPLICATION IN FIGURE 3 AT BASES 461-472.	33,34,38/37
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (USA), VOLUME 81, ISSUED OCTOBER 1984, MEYER ET AL., "PILUS GENES OF NEISSERIA GONORRHEAE: CHROMOSOMAL ORGANIZATION AND DNA SEQUENCE", PAGES 6110-6114, SEE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41
A	JOURNAL OF GENERAL MICROBIOLOGY, VOLUME 132, ISSUED 1986, TINSLEY ET AL., "VARIATION IN THE EXPRESSION OF PILI AND OUTER MEMBRANE PROTEIN BY NEISSERIA MENINGITIDIS DURING THE COURSE OF MENINGOCOCCAL INFECTION", PAGES 2483-2490, SEE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41

CORRECTED VERSION *

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07H 21/00, G01N 33/53

A1

(11) International Publication Number: WO 92/13871

(43) International Publication Date: 20 August 1992 (20.08.92)

(21) International Application Number: PCT/US92/00863

(22) International Filing Date: 31 January 1992 (31.01.92)

(30) Priority data:
648,781 31 January 1991 (31.01.91) US

(71) Applicant: WASHINGTON UNIVERSITY [US/US]; One Brookings Drive, St. Louis, MO 63130 (US).

(72) Inventors: NORMARK, Staffan ; 30 Ridgemoor Drive, Clayton, MO 63105 (US). JONSSON, Ann-Beth ; Stipendiegrand 4C, S-901 01 Umea (SE).

(74) Agents: MONROY, Gladys, H. et al.; Morrison & Foerster, 545 Middlefield Road, Suite 200, Menlo Park, CA 94025 (US).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, RU, SD, SE, SE (European patent), SN (OAPI patent), TD (OAPI patent), TG (OAPI patent).

Published
With international search report.

(54) Title: POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA

(57) Abstract

We have isolated and characterized a novel protein of pathogenic forms of *Neisseria*. We have also isolated and characterized genes which encode PilC, i.e., the *pilC* loci. Portions of the DNA sequences of the *pilC* genes are useful as probes to diagnose the presence of the relevant microorganisms containing type 4 pilin, for example, *Neisseria* in samples. These DNAs also make available polypeptide sequences of immunoreactive epitopes encoded within the loci, thus permitting the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or as components of vaccines. Antibodies, both monoclonal and purified polyclonal, directed against PilC epitopes are also useful for diagnostic test and as therapeutic agents for passive immunization. In addition, by utilizing probes derived from the DNA sequences, it is possible to isolate and sequence other portions of the *pilC* loci from species and strains of interest.

* (Referred to in PCT Gazette No. 24/1992, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
PL	Poland
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA

Technical Field

The invention relates to materials and methodologies for managing the spread of infections caused by microorganisms having type 4 pilin, for example, *Neisseria*. More specifically, it relates to polypeptides and antibodies useful in vaccines for the treatment of pathologic infections caused by these microorganisms. It also relates to polynucleotides useful for the recombinant production of these polypeptides. In addition, it relates to polypeptides, antibodies, and polynucleotides used for the detection of these strains.

Background Art

Type 4 pilins are expressed by several bacterial genuses, including *Neisseria, Moraxella, Bacteroides, and Pseudomonas*. Species within these genuses which have pathogenic members that express type 4 pilins are, for example, *N. goncrrhoeae, N. meningitidis, M. bovis, B. nodosus*, and *P. aeruginosa*. In addition, the Tcp pilin of *V. cholerae* is highly homologous to the type 4 pilins of other genuses.

The only known reservoir of the neisseriae is man. The genus includes two gram-negative species of pyogenic cocci that are pathogenic for man: the meningococcus (*Neisseria meningitidis*) and the gonococcus (*Neisseria gonorrheae*).

N. Meningitidis causes a variety of infections, most notably, meningitis and bacteremia. Meningococci can be divided into serologic groups on the basis of agglutination reactions with immune serum. The present classification includes groups A through Z. Clinically significant new groups encompass Y and W 135. The major groups are remarkably heterogeneous, but subclassification with additional serologic markers has been possible. Noncapsular antigens have provided the basis for dividing strains of groups into distinct types.

Meningococci cause either epidemic or sporadic disease, and historically, there has been a cyclic variation in the prevalence of meningococcal infection with peaks of increased frequency occurring every 8 to 12 years and lasting 4 to 6 years. The attack rate of meningococcal disease is highest for children between 6 months and 1 year. In the first half of this century, most epidemics of meningococcal disease in the United States were caused by group A organisms. In the past two decades, first group B then group C meningococci were responsible for outbreaks in both the military and civilian populations. Currently, group B is responsible for 50 to 55 percent of reported cases.

Gonorrhea, which is caused by *N. gonorrhea*, is an infection of columnar and transitional epithelium. This disease is the most common reportable communicable disease in the United States, and also has world-wide prevalence.

Although treatment of disease caused by gonococci and meningococci are often treated with antibiotics, these microorganisms often develop antibiotic resistance. Thus, prevention with vaccines is a preferable mode to contain the spread of infection. However, for a variety of reasons, including antigenic variation, the development of vaccines has been greatly hampered. For example, a vaccine which prevents gonorrhea is still lacking. In addition, although 56% of the causes of meningococcal disease are caused by serogroup B, an effective vaccine against this serogroup is also lacking.

N. gonorrheae and *N. meningitidis* are organisms completely adapted to the human host, having no other ecological niche. They have acquired a large arsenal of strategies to overcome the human host defense system.

The first step in infection with pathological forms of these *Neisseria* is adherence to target cells. It is thought that the pili of these microorganisms are a major virulence factor. For example, it is known that in the case of *N. gonorrheae*, piliated (P^+) variants attach much better to susceptible cells than non-piliated (P-) variants (Swanson, 1973; Pearce and Buchanan, 1978). Moreover, P+ variants, unlike P- variants, are able to establish an infection in human volunteers (Kellog et al, 1968).

Although the pilus protein elicits an immune response, so many antigenic variants exist and continue to develop that vaccines against the pilus protein are not highly effective.

Pilin is the major subunit of the pilus. Expression of pilin is controlled at the *pilE* locus.

Brief Description of the Invention

We have isolated and characterized a novel protein of pathogenic forms of *Neisseria*, PilC, that may be is associated with the pili of gonococci and meningococci. We have also isolated and characterized genes which encode PilC, i.e., the *pilC* loci.

Portions of the DNA sequences of the *pilC* genes are useful as probes to diagnose the presence of the relevant *Neisseria* in samples. These DNAs also make available polypeptide sequences of immunoreactive epitopes encoded within the loci, thus permitting the production of polypeptides which are useful as standards or reagents in diagnostic tests and/or as components of vaccines for microorganisms with type 4 pilin and containing one or more epitopes that are immunologically identifiable with an epitope encoded in *pilC* of *Neisseria*. Antibodies, both monoclonal and purified polyclonal, directed against PilC epitopes are also useful for diagnostic tests and as therapeutic agents for passive immunization. In addition, by utilizing probes derived from the DNA sequences, it is possible to isolate and sequence portions of the *pilC* loci from species and strains of interest.

Accordingly, one embodiment of the invention is a recombinant polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Another embodiment of the invention is a recombinant expression system comprising a polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*, wherein the polynucleotide is operably linked to a control sequence compatible with a desired host.

Yet another embodiment of the invention is purified polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Another embodiment of the invention is a recombinant polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Still another embodiment of the invention is a vaccine composition for the treatment of *Neisseria infection*, comprised of a pharmaceutically acceptable excipient and of an effective amount of a recombinant polypeptide, wherein the polypeptide is comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

Yet another embodiment of the invention is a composition comprised of purified polyclonal anti-PilC antibodies, wherein the PilC is of *Neisseria*.

An additional embodiment of the invention is a composition comprised of a monoclonal antibody directed against an immunoreactive epitope encoded in *pilC* of *Neisseria*.

Another embodiment of the invention is a method for producing antibodies to PilC of *Neisseria* comprising administering to an individual a composition comprised of an isolated immunogenic polypeptide containing a PilC epitope in an amount sufficient to produce an immune response.

Yet another embodiment of the invention is an oligomer capable of hybridizing to a sequence in *pilC* of *Neisseria*, wherein the oligomer is comprised of a *pilC* sequence complementary to at least about 6 contiguous nucleotides of *pilC*.

Still another embodiment of the invention is a process for detecting a *pilC* sequence in an analyte

strand, wherein the *pilC* sequence comprises a selected target region, the process comprising:

(a) providing a sample comprised of an analyte strand suspected of containing a selected target *pilC* sequence;

(b) providing an oligomer capable of hybridizing to the target *pilC* sequence, wherein the oligomer is comprised of a *pilC* targeting sequence complementary to at least about 6 contiguous nucleotides of *pilC*;

(c) incubating the sample of (a) with the oligomer of (b) under conditions which allow specific hybrid duplexes to form between the targeting sequence and the target sequence; and

(d) detecting hybrids formed between the target sequence, if any, and the oligomer.

Yet another embodiment of the invention is a recombinant polynucleotide comprising a DNA sequence of at least 8 contiguous nucleotides from *pilC*, wherein the *pilC* sequence is selected from the group of sequences shown in Figure 3, Figure 6, and Figure 7.

Another embodiment of the invention is a method of treating an individual for a *Neisseria* infection comprising administering to the individual antibodies produced according to claim 31, wherein the antibodies are administered in an amount effective to prevent the pathology of the infection.

Brief Description of the Drawings

Figure 1 is a genetic and physical map of *pilC* locus 1, showing the restriction enzyme sites.

Figure 2 shows the nucleotide sequence and the deduced amino acid sequence of the 5'-end of *pilC1*.

Figure 3 shows the nucleotide sequence of the sense strand of the *pilC1* gene.

Figure 4 shows the shows the nucleotide sequence of the sense strand of the *pilC1* gene and the amino acids encoded therein.

Figure 5 shows the nucleotide sequence of the sense strand of the *pilC1* gene, and the effect of frame shift on the putative gene products encoded therein.

Figure 6 shows the nucleotide sequences of PCR amplified fragments demonstrating a variation in the length of the G tract and sequence differences in the 5' region of the *pilC* genes.

Figure 7 shows the DNA sequence of the 3'-end of the *pilC2* fragment.

Figure 8 shows the *pilC2* fragment sequence, and the putative amino acids encoded therein.

Figure 9 shows a comparison of the analogous portions of *pilC2* (top) and *pilC1* (bottom) DNA sequences, and the putative amino acids encoded therein.

Modes for Carrying Out the Invention

The present invention provides polypeptides, antibodies, and polynucleotides which are useful for the detection and treatment of pathogenic microorganisms having type 4 pilin, for example, *Neisseria*, *Moraxella*, *Bacteroides*, and *Pseudomonas*.

We have discovered a polypeptide, PilC, which is present in *N. gonorrheae*. This polypeptide is a 110 kd protein that is closely associated with the pili of the microorganism. Most strains of *N. gonorrheae* carry two copies of the corresponding genes which encode the polypeptide(s); these genes have been denoted *pilC*. Expression from the *pilC* loci is regulated by frequent frameshift mutations within a run of G residues in the

region encoding the signal peptide. The two *pilC* genes of *N. gonorrheae* are not identical. Hence, alternate expression from either the *pilC1* or the *pilC2* loci gives rise to two different forms of PilC. Among nonpiliated (P^-) descendants from P^+ clones, clones were found that expressed pilin but not PilC. All P^+ revertants from such $PilC^-$ non-piliated clones have regained expression of PilC. Hence, phase variation of gonococcal pili can be caused by frameshift mutations in *pilC*. Transposon inactivation of the expressed *pilC2* copy resulted in a nonpiliated, pilin producing revertible phenotype. It appears, therefore, that PilC is required for assembly of pilin subunits into a polymerized pilus fiber in *N. gonorrheae*.

We have cloned and isolated gene, *pilC1*, from *N. gonorrheae*. In addition, by comparison of this gene sequence with a related sequence, we have cloned a fragment of the *pil2* gene. Moreover, using polynucleotide probes derived from isolated *pilC1* and PCR amplification, we have detected two possible variants of a *pilC* gene in *N. meningitidis*. The sequences of *pilC* reported herein appear to be novel, in that there are no reported counterparts in Genbank, and no significant homologies were found with any available sequences in that data base.

The useful materials and processes of the present invention are made possible by the provision of the sequences of the *pilC* genes from *N. gonorrheae* and from *N. meningitidis*. Information present in the sequences of the *pilC* genes allows for the design of polypeptides which may be useful as vaccines for treatment of pathogenic *Neisseria*, as diagnostic tools for the detection of these microorganisms, and as agents for the preparation of antibodies to PilC. In addition,

this information allows for the design of polynucleotides for the recombinant production of the polypeptides derived from PilC, and for the design of oligomers which are useful as probes and primers for the detection and amplification of target regions of *pilC*.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Maniatis, Fitsch & Sambrook, MOLECULAR CLONING; A LABORATORY MANUAL, Second Edition (1989); DNA CLONING, VOLUMES I AND II (D.N Glover ed. 1985); OLIGONUCLEOTIDE SYNTHESIS (M.J. Gait ed, 1984); NUCLEIC ACID HYBRIDIZATION (B.D. Hames & S.J. Higgins eds. 1984); TRANSCRIPTION AND TRANSLATION (B.D. Hames & S.J. Higgins eds. 1984); ANIMAL CELL CULTURE (R.I. Freshney ed. 1986); IMMOBILIZED CELLS AND ENZYMES (IRL Press, 1986); B. Perbal, A PRACTICAL GUIDE TO MOLECULAR CLONING (1984); the series, METHODS IN ENZYMOLOGY (Academic Press, Inc.); GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory), Methods in Enzymology Vol. 154 and Vol. 155 (Wu and Grossman, and Wu, eds., respectively), Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London), Scopes, (1987), PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.), and HANDBOOK OF EXPERIMENTAL IMMUNOLOGY, VOLUMES I-IV (D.M. Weir and C. C. Blackwell eds 1986). All patents, patent applications, and publications mentioned herein, both supra and infra, are hereby incorporated herein by reference.

As used herein, a polynucleotide "derived from" a designated sequence refers to a polynucleotide sequence which is comprised of a sequence of approximately at least about 6 nucleotides, preferably at least about 8 nucleotides, more preferably at least about 10-12 nucleotides, and even more preferably at least about 15-20 nucleotides corresponding to a region of the designated nucleotide sequence. "Corresponding" means homologous to or complementary to the designated sequence. Regions from which typical polynucleotide sequences may be "derived" include but are not limited to, for example, regions encoding specific epitopes, as well as non-transcribed and/or non-translated regions.

The derived polynucleotide is not necessarily physically derived from the nucleotide sequence shown, but may be generated in any manner, including for example, chemical synthesis or DNA replication or reverse transcription or transcription. In addition, combinations of regions corresponding to that of the designated sequence may be modified in ways known in the art to be consistent with an intended use.

Similarly, a polypeptide or amino acid sequence "derived from" a designated nucleic acid sequence refers to a polypeptide having an amino acid sequence identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 3-5 amino acids, and more preferably at least 8-10 amino acids, and even more preferably at least 11-15 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence. This terminology also includes a polypeptide expressed from a designated nucleic acid sequence.

A recombinant or derived polypeptide is not necessarily translated from a designated nucleic acid

sequence. It may be generated in any manner, including for example, chemical synthesis, or expression of a recombinant expression system, or isolation from a microorganism. A recombinant or derived polypeptide may include one or more analogs of amino acids or unnatural amino acids in its sequence. Methods of inserting analogs of amino acids into a sequence are known in the art. It also may include one or more labels, which are known to those of skill in the art.

The term "recombinant polynucleotide" as used herein intends a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, (3) does not occur in nature, or (4) is not in the form of a library.

The term "polynucleotide" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications, for example, labels which are known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example proteins (including for e.g., nucleases, toxins, antibodies, signal peptides,

poly-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide.

The term "purified polynucleotide" refers to a polynucleotide which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90% of polypeptides with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides from bacteria are known in the art, and include for example, disruption of the bacteria with a chaotropic agent, differential extraction and separation of the polynucleotide(s) and polypeptides by ion-exchange chromatography, affinity chromatography, and sedimentation according to density.

The term "purified polypeptide" refers to a polypeptide or fragment thereof which is essentially free, i.e., contains less than about 50%, preferably less than about 70%, and even more preferably less than about 90%, of cellular components with which the polypeptide is naturally associated. Techniques for purifying polypeptides are known in the art, and examples of these techniques are discussed infra.

"Recombinant host cells", "host cells", "cells", "cell lines", "cell cultures", and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be, or have been, used as recipients for recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in

morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

A "replicon" is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc. that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

A "vector" is a replicon in which another polynucleotide segment is attached, so as to bring about the replication and/or expression of the attached segment.

"Control sequence" refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and terminators; in eukaryotes, generally, such control sequences include promoters, terminators and, in some instances, enhancers. The term "control sequences" is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

An "open reading frame" (ORF) is a region of a polynucleotide sequence which encodes a polypeptide; this

region may represent a portion of a coding sequence or a total coding sequence.

A "coding sequence" is a polynucleotide sequence which is transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus. A coding sequence can include, but is not limited to mRNA, cDNA, and recombinant polynucleotide sequences.

"Immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptides(s) which are also present in the designated polypeptide(s). Immunological identity may be determined by antibody binding and/or competition in binding; these techniques are known to those of average skill in the art, and are also illustrated infra.

As used herein, "epitope" refers to an antigenic determinant of a polypeptide. An epitope could comprise 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 5 such amino acids, and more usually, consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance.

A polypeptide is "immunoreactive" when it is "immunologically reactive" with an antibody, i.e., when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly by the kinetics of antibody

binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an epitope against which the antibody is directed. The techniques for determining whether a polypeptide is immunologically reactive with an antibody are known in the art. An "immunoreactive" polypeptide may also be "immunogenic". As used herein, the term "immunogenic polypeptide" is a polypeptide that elicits a cellular and/or humoral immune response, whether alone or linked to a carrier in the presence or absence of an adjuvant.

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one antibody combining site. An "antibody combining site" or "binding domain" is formed from the folding of variable domains of an antibody molecule(s) to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows an immunological reaction with the antigen. An antibody combining site may be formed from a heavy and/or a light chain domain (VH and VL, respectively), which form hypervariable loops which contribute to antigen binding. The term "antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, altered antibodies, univalent antibodies, the Fab proteins, and single domain antibodies.

As used herein, a "single domain antibody" (dAb) is an antibody which is comprised of an VH domain, which reacts immunologically with a designated antigen. A dAB does not contain a VL domain, but may contain other antigen binding domains known to exist in antibodies, for example, the kappa and lambda domains. Methods for preparing dABs are known in the art. See, for example, Ward et al. (1989).

Antibodies may also be comprised of VH and VL domains, as well as other known antigen binding domains. Examples of these types of antibodies and methods for their preparation are known in the art (see, e.g., U.S. Patent No. 4,816,467, which is incorporated herein by reference), and include the following. For example, "vertebrate antibodies" refers to antibodies which are tetramers or aggregates thereof, comprising light and heavy chains which are usually aggregated in a "Y" configuration and which may or may not have covalent linkages between the chains. In vertebrate antibodies, the amino acid sequences of all the chains of a particular antibody are homologous with the chains found in one antibody produced by the lymphocyte which produces that antibody *in situ*, or *in vitro* (for example, in hybridomas). Vertebrate antibodies typically include native antibodies, for example, purified polyclonal antibodies and monoclonal antibodies. Examples of the methods for the preparation of these antibodies are described infra.

"Hybrid antibodies" are antibodies wherein one pair of heavy and light chains is homologous to those in a first antibody, while the other pair of heavy and light chains is homologous to those in a different second antibody. Typically, each of these two pairs will bind different epitopes, particularly on different antigens. This results in the property of "divalence", i.e., the ability to bind two antigens simultaneously. Such hybrids may also be formed using chimeric chains, as set forth below.

"Chimeric antibodies", are antibodies in which the heavy and/or light chains are fusion proteins. Typically the constant domain of the chains is from one particular species and/or class, and the variable domains

are from a different species and/or class. Also included is any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be differing classes, or different species of origin, and whether or not the fusion point is at the variable/constant boundary. Thus, it is possible to produce antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation, or to make other improvements in properties possessed by a particular constant region.

Another example is "altered antibodies", which refers to antibodies in which the naturally occurring amino acid sequence in a vertebrate antibody has been varied. Utilizing recombinant DNA techniques, antibodies can be redesigned to obtain desired characteristics. The possible variations are many, and range from the changing of one or more amino acids to the complete redesign of a region, for example, the constant region. Changes in the constant region, in general, to attain desired cellular process characteristics, e.g., changes in complement fixation, interaction with membranes, and other effector functions. Changes in the variable region may be made to alter antigen binding characteristics. The antibody may also be engineered to aid the specific delivery of a molecule or substance to a specific cell or tissue site. The desired alterations may be made by known techniques in molecular biology, e.g., recombinant techniques, site directed mutagenesis, etc.

Yet another example are "univalent antibodies", which are aggregates comprised of a heavy chain/light chain dimer bound to the Fc (i.e., constant) region of a second heavy chain. This type of antibody escapes antigenic modulation. See, e.g., Glennie et al. (1982).

Included also within the definition of antibodies are "Fab" fragments of antibodies. The "Fab" region refers to those portions of the heavy and light chains which are roughly equivalent, or analogous, to the sequences which comprise the branch portion of the heavy and light chains, and which have been shown to exhibit immunological binding to a specified antigen, but which lack the effector Fc portion . "Fab" includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers containing the 2H and 2L chains (referred to as $F(ab)_2$), which are capable of selectively reacting with a designated antigen or antigen family. "Fab" antibodies may be divided into subsets analogous to those described above, i.e, "vertebrate Fab", "hybrid Fab", "chimeric Fab", and "altered Fab". Methods of producing "Fab" fragments of antibodies are known within the art and include, for example, proteolysis, and synthesis by recombinant techniques.

The term "polypeptide" refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages, as well as

other modifications known in the art, both naturally occurring and non-naturally occurring.

"Transformation", as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for the insertion, for example, direct uptake, transduction, f-mating or electroporation. The exogenous polynucleotide may be maintained as a non-integrated vector, for example, a plasmid, or alternatively, may be integrated into the host genome.

"Treatment" as used herein refers to prophylaxis and/or therapy.

By "immunogenic" is meant an agent used to stimulate the immune system of a living organism, so that one or more functions of the immune system are increased and directed towards the immunogenic agent. Immunogenic agents include vaccines. Immunogenic agents can be used in the production of antibodies, both isolated polyclonal antibodies and monoclonal antibodies, using techniques known in the art.

By vaccine is meant an agent used to stimulate the immune system of a living organism so that protection against or amelioration of future harm is provided. Immunization refers to the process of inducing an increased level of antibodies and/or cellular immune response in which T-lymphocytes respond by killing the pathogen and/or activate other cells involved in the immune response pathway. The antibodies produced as a result of immunization may belong to any of the immunological classes, such as immunoglobulins A, D, E, G or M.

An "individual", as used herein, refers to vertebrates, particularly members of the mammalian

species, and includes but is not limited to domestic animals, sports animals, and primates, including humans.

As used herein, the "sense strand" of a nucleic acid contains the sequence that has sequence homology to that of mRNA. The "anti-sense strand" contains a sequence which is complementary to that of the "sense strand".

As used herein, the term "probe" refers to a polynucleotide which forms a hybrid structure with a sequence in a target region, due to complementarity of at least one sequence in the probe with a sequence in the target region. The polynucleotide regions of probes may be composed of DNA, and/or RNA, and/or synthetic nucleotide analogs.

As used herein, the term "target region" refers to a region of the nucleic acid which is to be amplified and/or detected. The term "target sequence refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

The term "primer" as used herein refers to an oligomer which is capable of acting as a point of initiation of synthesis of a polynucleotide strand when placed under appropriate conditions. The primer will be completely or substantially complementary to a region of the polynucleotide strand to be copied. Thus, under conditions conducive to hybridization, the primer will anneal to the complementary region of the analyte strand. Upon addition of suitable reactants, (e.g., a polymerase, nucleotide triphosphates, and the like), the primer is extended by the polymerizing agent to form a copy of the analyte strand. The primer may be single-stranded, or alternatively may be partially or fully double-stranded.

The terms "analyte polynucleotide" and "analyte strand" refer to a single- or double-stranded nucleic

acid molecule which is suspected of containing a target sequence, and which may be present in a biological sample.

As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs, and also samples of *in vitro* cell culture constituents (including but not limited to conditioned medium resulting from the growth of cells in cell culture medium, putatively virally infected cells, recombinant cells, and cell components).

As used herein, the term "oligomer" refers to primers and to probes. The term oligomer does not connote the size of the molecule. However, typically oligomers are no greater than 1000 nucleotides, more typically are no greater than 500 nucleotides, even more typically are no greater than 250 nucleotides; they may be no greater than 100 nucleotides, and may be no greater than 75 nucleotides, and also may be no greater than 50 nucleotides in length.

The term "coupled" as used herein refers to attachment by covalent bonds or by strong non-covalent interactions (e.g., hydrophobic interactions, hydrogen bonds, etc.). Covalent bonds may be, for example, ester, ether, phosphoester, amide, peptide, imide, carbon-sulfur bonds, carbon-phosphorus bonds, and the like.

The term "support" refers to any solid or semi-solid surface to which a desired polypeptide or polynucleotide may be anchored. Suitable supports include glass, plastic, metal, polymer gels, and the

like, and may take the form of beads, wells, dipsticks, membranes, and the like.

The term "label" as used herein refers to any atom or moiety which can be used to provide a detectable (preferably quantifiable) signal, and which can be attached to a polynucleotide or polypeptide.

The term "type 4 pilin" as used herein refer to pilins that contain a conserved amino terminal hydrophobic domain beginning with an amino-terminal phenylalanine that is methylated upon processing and secretion of the pilin. Another characteristic feature of type 4 pilins is that in the propilin form they contain similar six- or seven-amino acid long leader peptides, which are much shorter than typical signal sequences. Type 4 pilins are expressed by several bacterial genuses, including *Neisseria, Moraxella, Bacteroides, and Pseudomonas*. Species within these genuses which express type 4 pilins are, for example, *N. gonorrhoeae, N. meningitidis, M. bovis, B. nodosus*, and *P. aeruginosa*. As used herein, the term "type 4 pilin" also includes the Tcp pilin of *Vibrio*, (for example, V. cholerae), that is highly homologous to the type 4 pilins of other genuses. Tcp pilin contains the characteristic amino-terminal hydrophobic domain as well as having a modified N-terminal amino acid that in this case may be a modified methionine because the Tcp pilin gene encodes a methionine residue at the position where all the others encode a phenylalanine. Precursor TcpA contains a much longer leader sequence than typical type 4 propilins but retains homology in the region surrounding the processing site.

The term "*pilC*" as used herein refers to a gene encoding a polypeptide involved in the assembly of type 4 pilin, which may also be required for attachment of the

pilin, and that is comprised of an epitope that is immunologically identifiable with an epitope in PilC of *N. gonorrhae* or *N. meningitidis*. Included within this term is any homologous region from *Vibrio*, *tcpC*.

As used herein the term "PilC" refers to a polypeptide encoded within *pilC*, and includes TcpC of *Vibrio*.

The description of the method to retrieve the DNA sequences is mostly of historical interest. The resultant sequences (and their complements) are provided herein, and the sequences, or any portion thereof, could be prepared using synthetic methods, or by a combination of synthetic methods with retrieval of partial sequences using methods similar to those described herein.

The description infra, of "walking" the genome by isolating overlapping DNA sequences from the *N. gonorrheae* lambda gt-11 library and from an EMBL3 library provides one method by which DNAs corresponding to the *pilC* genomes from, inter alia, *N. gonorrheae* and *N. meningitidis*, respectively, may be isolated. However, given the information provided herein, other methods for isolating *pilC* DNAs from these species, as well as from species of other genuses which have type 4 pilin are obvious to one of skill in the art.

Characterization of the genes of the *pilC* loci has provided information on the polypeptides encoded therein, and on the control of their expression. Even though Type 4 pili have been extensively studied in several laboratories, little is known about their assembly. The presence of a specific assembly machinery for this class of pili is evident from the fact that the pilin gene of *B. nodosus* and *M. bovis* can be properly processed and assembled into a pilus in *P. aeruginosa* but not in *E. coli* (Ellerman et al., 1986; Mattick et al.,

1987; Beard et al., 1990). Furthermore, the recent genetic characterization of TCP pili of *Vibrio cholerae* has revealed that a number of closely linked genes are required for pilin processing and assembly into a structure (Taylor et al., 1988). The TCP pilin does not carry an N-methylphenylalanine but its primary sequence is highly homologous to the Type 4 class of pilins.

The *N. gonorrhoeae* pilus facilitates adherence of the bacterium to a number of eukaryotic cell types (Watt et al., 1980) and is thought to play a role in bacterial interaction with neutrophils (Fischer and Rest, 1988). The pilin is encoded from one or two *pilE* loci (Meyer et al., 1984; Swanson et al., 1986) which most likely each form a monocistronic operon. Hence, there have been no suggestions that genes closely linked to *pilE* are involved in pilus assembly. A dispersed location of genes involved in gonococcal pilus assembly as well as the rapid occurrence of nonpiliated variants generated via recombination with pilin sequences from silent loci, *pilS*, have made it extremely difficult to identify putative assembly genes for gonococcal pili.

The PilC protein described herein is a protein encoded within a *pilC* or equivalent (for example, *tcpC*) locus or gene. In *N. gonorrhoeae* MS11 and most other gonococcal strains the PilC protein is expressed in small amounts. It is the only protein that is enriched in highly purified preparations of MS11 pili. PilC was not released from a nonpilated MS11 (P^-n) variant using the same procedure suggesting that this protein interacts with the polymerized pilus fiber.

DNA sequence analysis of the cloned *pilC1* gene revealed one long open reading frame that was out of frame with its putative AUG initiation codon and 5' end encoding the signal peptide. Minute amounts of PilC were

expressed in *E. coli* from pABJ04. Gel purified PilC from MS11 contained a lysine residue in position four, whereas *pilC1* had a glutamine codon at this position. A lysine codon was, however, found at position four in a number of PCR amplified 5' *pilC* fragments suggesting that these fragments represent the 5' end of *pilC2*, which then must be ON in MS11. The finding that a miniTnCm insertion in *pilC2* abolished PilC expression, whereas insertional inactivation of *pilC1* did not abolish PilC expression further argues that *pilC1* is translationally out of frame and *pilC2* translationally in frame in the MS11 variant we are studying.

PCR amplified fragments of *pilC1* and *pilC2* in MS11 differed in the number of G residues found in the G tract. Only 11 or 12 Gs were found in *pilC1* clones (which would both generate an OFF phenotype) while 12 or 13 Gs were found among *pilC2* specific clones. Since *pilC2* is the expressed gene in the MS11 variant under study, we believe that this variant carries 13 Gs in *pilC2* and 12 Gs in *pilC1*. The frequency of frameshift mutations in each locus is not known. However, the lack of 13 Gs among *pilC1* specific fragments and the lack of 14 Gs among *pilC2* specific fragments suggests that a deletion of one G residue occurs at a higher frequency than the insertion of one G residue. We had expected to find amplified fragments from *N. gonorrhoeae* containing 10 G residues in the G tract, but found none in the 48 clones sequenced. If only one G is added or deleted in each mutational event, the frequency of G tracts with 10 residues should be low if G tracts normally are 12 or 13 bp long.

Frameshifting in *pilC1* also occurred in *E. coli*. In this case, however, two variants with 10 residues were found out of 12 clones sequenced. It may

therefore be that there is a selection against in frame variants with 10 Gs in *N. gonorrhoeae*. A change from five glycines to four in the signal peptide may for example have an effect on the physical properties of the precursor form of PilC such that the signal peptide is not cleaved off. *E. coli* strain AA10 is *recA*. Therefore, frameshift mutations in the G tract of *pilC* occurs independent of the RecA protein.

Translational frameshifting has been shown to regulate phase and antigenic variation of the gonococcal opacity protein PII that is encoded by a number of opa loci showing sequence variations. In this system a number of pentameric CTCTT repeats are present in the region encoding the signal peptide (Stern et al. 1986). Variation in the number of repeats is independent of *recA* in *N. gonorrhoeae* as well as in *E. coli* (Murphy et al., 1989). Variation in the expression of lipopolysaccharide epitopes in *Haemophilus influenzae* was recently explained by translational frameshifting created by alterations in the number of CAAT repeats occurring in the 5' end of *licA* (Weiser et al., 1989). In *Bordetella pertussis* frameshift mutations in the regulatory *vir* locus occur in a run of C residues positioned internally in the gene (Stibitz et al., 1989). The C tract was in this case varying from 6 (in frame) to 7 residues (out of frame). It is not known if this frameshift mutation is programed or not. The pilin gene of *Bordetella pertussis* was recently shown to be preceded by a stretch of Cs. Frequent mutations affecting the length of this C tract influenced the transcriptional activity of the pilin gene (Willems et al., 1990).

Variation in the number of the CTCTT repeats in opa genes was recently suggested to be due to recombination-independent slipped strand mispairing

(Murphy et al., 1989). Mispairing is thought to occur between strands subjected to local denaturation and should preferentially occur during replication. A number of unusual DNA structures (cruciform, Z form, H form) have been shown to form *in vitro* within a variety of specific DNA sequences. Under normal conditions the B form is the most favorable thermodynamically (Frank-Kamenetskii and Vologodskii, 1984). Transition to alternative conformations requires specific external conditions, supercoiling being the most physiologic. Single stranded $(dG)_n$ and $(dC)_n$ strands renature more slowly than complementary strands with arbitrary sequences, and methylation experiments suggest that a poly dG chain may form a hairpin-like structure stabilized by G-G bp (Panyutin et al., 1990).

Four variant sequences differing outside the G tract were obtained by PCR amplification of the 5' end of *pilC* from four *N. gonorrhoeae* strains. The region 5' of the G tract was invariant, as was the 3' end of the amplified region. All variation was confined to a region located 3' of the G tract. At least some of these sequence variations can be explained by mismatch pairing events. Thus, the addition of four nucleotides distal to the G tract in variant sequence 4 is possible to explain by a two step mispairing event occurring within variant sequence 2. Slip strand mispairing between the two CA residues in -GGCGCAGGCGCA- would yield -GGCGCAGGCGCAGGCGCA-. A second mispairing event occurring between the two C-residues at positions 3 and 5 gives rise to the sequence -GGCAGGCGCAGGCGCA- present in variant 4. It may therefore be that a sequence close to a poly(G) tract is prone to slipped strand mispairing.

Gonococcal pilus phase variation is associated with an altered nucleotide sequence of *pilE* via

recombinations with silent *pilS* sequences (Haas and Meyer, 1986, Swanson et al., 1986). An irreversible switch OFF in pilus expression results from deletions of the 5' coding and control regions of the *pilE* locus (Swanson et al., 1985). Reversible gonococcal pilus phase variation is associated with nucleotide changes in *pilE* resulting in an altered pilin product. It has been suggested that the pilins of these variants are assembly defective (Bergstrom et al., 1986; Swanson et al., 1986; Hill et al., 1990). Here we present evidence that switch OFF and ON of PilC expression causes pilus expression to phase vary. Five out of five P^-, pilin producing descendants from $MS11_{mk}$ (P^+, $PilC^+$) that expressed pilin did not express PilC. All tested P^+ revertants from the five P^-, $PilC^-$ variants had regained expression of PilC. The pilin of one nonpiliated PilC OFF-switcher (variant 8) differed by eight amino acids from that of the parent. The fact that one piliated $PilC^+$ backswitcher (8:1) expressed a pilin identical in sequence to the nonpiliated variant (8) strongly suggests that the regained expression of pili is due to an ON-switch in PilC expression. The above results also imply that the nonpiliated phenotype of variant 8 is not due to the alterations in the pilin relative to the parental strain but to an OFF-switch of PilC. The finding that mTnCm insertions resulted in P^+ colonies when inserted into *pilC1* and P^- colonies when inserted into the actively expressing *pilC2* locus offers further evidence that PilC is essential for the biogenesis of gonococcal pili. P^-, *pilC2*::mTnCm-12 insertion mutants reverted to P^+ colony morphology at a low frequency. These revertants most likely represent frameshifting mutants in *pilC1* resulting in expression of PilC from this locus. A double mutant in *pilC1* and *pilC2* was stably nonpiliated, expressed

pilin, but expressed amounts of pilin that did not express any pili when examined by transmission electron microscopy. It is therefore believed that out of frame mutations of both *pilC1* and *pilC2* will abolish pili formation.

At this stage we cannot exclude the possibility that some PilC$^-$ variants from MSII$_{mk}$(P$^+$,PilC$^-$) are generated by transformation of *pilC1* sequences and homologous recombination with *pilC2* thus generating variants with two *pilC1* 5' ends at both *pilC* loci. PilC$^+$ revertants from PilC$^-$ clones must, however, all be due to frameshift mutations in either *pilC1* or *pilC2*.

We propose that PilC forms an outer membrane pore or assembly center enabling the pilin subunits to be assembled and translocated across the outer membrane analogous to the proposed function of the high molecular weight proteins required for the assembly of enterobacterial pili (the latter of which is discussed in Norgren et al., 1987). Alternatively, PilC may act as an initiator for polymerization. In the latter case PilC would be expected to be located at the tip of the polymerized pilus.

It is possible that the alternate expression of PilC from two structurally different *pilC* loci is yet another example of antigenic variation in *Neisseria gonorrhoeae*. It is, however, possible that this variation could have functional implications as well. Each class of *E. coli* pili utilizes a different outer membrane pore/assembly protein. Hence, pilin subunits and/or periplasmic chaperone complexes may specifically interact with an exposed region of the protein allowing polymerization of pilus subunit proteins. The repertoire of antigenic variants of gonococcal pilins is vast (Hagblom et al., 1985). It may be that only certain

pilin variants are assembled via PilC1 and PilC2 respectively. Alternatively, if PilC acts as an initiator it could also possess other properties such as being involved in Pilus mediated attachment.

In one embodiment of the invention, immunogenically active polypeptides encoded within *pilC* are prepared. The availability of *pilC* DNA sequences, either those isolated by utilizing the DNA sequences described in the Examples, or nucleotide sequences derived therefrom (including segments and modifications of the sequence), permits the construction of expression vectors encoding immunologically reactive regions of the polypeptide encoded in either strand. Immunological reactivity may be determined by immunoassay using antibodies raised to PilC. Fragments encoding the desired polypeptides are derived from the DNA clones using conventional restriction digestion or by synthetic methods, and are ligated into vectors which may, for example, contain portions of fusion sequences such as beta-galactosidase or superoxide dismutase (SOD), preferably SOD. Methods and vectors which are useful for the production of polypeptides which contain fusion sequences of SOD are described in European Patent Office Publication number 0196056, published October 1, 1986. Any desired portion of the *pilC* DNA containing an open reading frame, in either sense strand, can be obtained as a recombinant polypeptide, such as a mature or fusion protein; alternatively, a polypeptide encoded in the DNA can be provided by chemical synthesis.

The DNA encoding the desired polypeptide, whether in fused or mature form, and whether or not containing a signal sequence to permit secretion, may be ligated into expression vectors suitable for any convenient host. Both eukaryotic and prokaryotic host

systems are presently used in forming recombinant polypeptides, and a summary of some of the more common control systems and host cell lines is given infra. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification may be by techniques known in the art, for example, differential extraction, salt fractionation, chromatography on ion exchange resins, affinity chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins. Such polypeptides can be used as diagnostics, or those which give rise to neutralizing antibodies may be formulated into vaccines. Antibodies raised against these polypeptides can also be used as diagnostics, or for passive immunotherapy.

The PilC antigens may also be isolated from meningococci and from gonococci. The bacteria may be grown by conditions known in the art, some of which are described infra. In addition, a method for isolating PilC from gonococci is described infra.

In another embodiment of the invention, the immunoreactive polypeptides may be conjugated with carrier. An antigenic region of a polypeptide is generally relatively small--typically 8 to 10 amino acids or less in length. Fragments of as few as 5 amino acids may characterize an antigenic region. These segments may correspond to regions of PilC antigen. Accordingly, using the DNAs of *pilC* as a basis, DNAs encoding short segments of PilC polypeptides can be expressed recombinantly either as fusion proteins, or as isolated polypeptides. In addition, short amino acid sequences can be conveniently obtained by chemical synthesis. In instances wherein the synthesized polypeptide is correctly configured so as to provide the correct

epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridyl-thio)propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue.) These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun. Rev. (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl groups can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized

macromolecules such as proteins; polysaccharides, such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, and the like; amino acid copolymers; and inactive virus particles or attenuated bacteria of other strains, for example, those of *Salmonella*. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

In addition to full-length PilC proteins, polypeptides comprising truncated PilC amino acid sequences encoding at least one immunologically reactive epitope are useful immunological reagents. For example, polypeptides comprising such truncated sequences can be used as reagents in an immunoassay. These polypeptides also are candidate subunit antigens in compositions for antiserum production or vaccines. While these truncated sequences can be produced by various known treatments of native bacterial protein, it is generally preferred to make synthetic or recombinant polypeptides comprising a PilC sequence. Polypeptides comprising these truncated PilC sequences can be made up entirely of PilC sequences (one or more epitopes, either contiguous or noncontiguous), or PilC sequences and heterologous sequences in a fusion protein. Useful heterologous sequences include sequences that provide for secretion from a recombinant host, enhance the immunological reactivity of the PilC epitope(s), or facilitate the coupling of the polypeptide to an immunoassay support or a vaccine carrier. See, e.g., EPO Pub. No. 116,201; U.S. Pat. No. 4,722,840; EPO Pub. No. 259,149; U.S. Pat. No. 4,629,783, the disclosures of which are incorporated herein by reference.

The size of polypeptides comprising the truncated PilC sequences can vary widely, the minimum size being a sequence of sufficient size to provide an immunologically reactive PilC epitope, while the maximum size is not critical. For convenience, the maximum size usually is not substantially greater than that required to provide the desired PilC epitopes and function(s) of the heterologous sequence, if any. Typically, the truncated PilC amino acid sequence will range from about 5 to about 100 amino acids in length. More typically, however, the PilC sequence will be a maximum of about 50 amino acids in length, preferably a maximum of about 30 amino acids. It is usually desirable to select PilC sequences of at least about 10, 12 or 15 amino acids, up to a maximum of about 20 or 25 amino acids.

Truncated PilC amino acid sequences comprising epitopes can be identified in a number of ways. For example, the entire PilC protein sequence can be screened by preparing a series of short peptides that together span the entire protein sequence. By starting with, for example, 100mer polypeptides, it would be routine to test each polypeptide for the presence of epitope(s) showing a desired reactivity, and then testing progressively smaller and overlapping fragments from an identified 100mer to map the epitope of interest. Screening such peptides in an immunoassay is within the skill of the art. It is also known to carry out a computer analysis of a protein sequence to identify potential epitopes, and then prepare oligopeptides comprising the identified regions for screening.

In another embodiment of the invention, the immunogenicity of the epitopes of PilC may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins

such as, for example, that associated with hepatitis B surface antigen. See, e.g., U.S. Pat. No. 4,722,840. Constructs wherein the PilC epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the PilC epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide. Thus, particles constructed from particle forming protein which include PilC sequences are immunogenic with respect to the microorganism encoding the PilC epitope (for example, *Neisseria, Vibrio, Moraxella, Bacteroides, or Pseudomonas*) and HBV.

Hepatitis surface antigen (HBSAg) has been shown to be formed and assembled into particles in *S. cerevisiae* (Valenzuela et al. (1982)), as well as in, for example, mammalian cells (Valenzuela, P., et al. (1984)). The formation of such particles has been shown to enhance the immunogenicity of the monomer subunit. The constructs may also include the immunodominant epitope of HBSAg, comprising the 55 amino acids of the presurface (pre-S) region. Neurath et al. (1984). Constructs of the pre-S-HBSAg particle expressible in yeast are disclosed in EPO 174,444, published March 19, 1986; hybrids including heterologous viral sequences for yeast expression are disclosed in EPO 175,261, published March 26, 1966. These constructs may also be expressed in mammalian cells such as Chinese hamster ovary (CHO) cells using an SV40-dihydrofolate reductase vector (Michelle et al. (1984)).

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding a PilC epitope. In this replacement, regions which are not required to mediate the aggregation of the

units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the PilC epitope.

In another embodiment of the invention, the immunoreactive polypeptides encoded in *pilC* are prepared into vaccines. Vaccines may be prepared from one or more immunogenic polypeptides derived from *pilC*. If recombinant, these polypeptides may be expressed in various host cells (e.g., bacteria, yeast, insect, or mammalian cells), or alternatively may be isolated from the bacterial preparations. In addition to the above, it is also possible to prepare live vaccines of attenuated microorganisms which express one or more recombinant polypeptides derived from the *pilC* gene. Suitable attenuated microorganisms are known in the art and include, for example, viruses (e.g., vaccinia virus (see Brown et al. (1986)), as well as bacteria.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples

of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing a PilC immunoreactive sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1%-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10%-95% of active ingredient, preferably 25%-70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 micrograms to 250 micrograms of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgment of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgment of the practitioner.

In addition, the vaccine containing the immunogenic antigen(s) derived from *pilC* may be administered in conjunction with other immunoregulatory agents, for example, immune globulins.

Another embodiment of the invention are antibodies which react immunologically with PilC epitopes. The immunogenic polypeptides prepared as described above are used to produce antibodies, including polyclonal and monoclonal. If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunized with an immunogenic polypeptide bearing a PilC epitope(s). Serum from the immunized animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to a PilC epitope (i.e., an epitope encoded within *pilC*) contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987).

Monoclonal antibodies directed against PilC epitopes can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al. (1980); Hammerling et al. (1981); Kennett et al. (1980); see also, U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,466,917; 4,472,500; 4,491,632; and 4,493,890. Panels of monoclonal antibodies produced against PilC epitopes can be screened

for various properties; i.e., for isotype, epitope affinity, etc.

Antibodies, both monoclonal and polyclonal, which are directed against PilC epitopes are particularly useful in diagnosis, and those which are neutralizing are useful in passive immunotherapy. Methods for introducing antibodies into an individual to accomplish passive immunotherapy are known in the art. In addition, monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies.

Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the antigen of the infectious agent against which protection is desired. See, for example, Nisonoff, A., et al. (1981) and Dreesman et al. (1985). Techniques for raising anti-idiotype antibodies are known in the art. See, for example, Grzych (1985), MacNamara et al. (1984), and Uytdehaag et al. (1985). These anti-idiotype antibodies may also be useful for treatment, vaccination and/or diagnosis of the relevant microorganism encoding the antigen of interest, (for example, *Neisseria*, *Pseudomonas*, *Moraxella*, *Bacteroides*, or *Vibrio*) as well as for an elucidation of the immunogenic regions of PilC.

Another embodiment of the invention concerns immunoassays and diagnostic kits. The polypeptides which contain epitopes encoded in *pilC* which are immunoreactive with anti-PilC antibodies in biological samples are useful in immunoassays to detect presence of anti-PilC antibodies, or the presence of the relevant microorganism or its antigens in biological samples. Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. The immunoassay will utilize a polypeptide comprised of at least one epitope derived from PilC or encoded in *pilC*. In one embodiment,

the immunoassay uses a combination of epitopes including the one derived from PilC or encoded in *pilC*. These epitopes may be derived from the same or from different polypeptides, and may be in separate recombinant or natural polypeptides, or together in the same recombinant polypeptides. An immunoassay may use, for example, a monoclonal antibody directed towards an epitope(s), a combination of monoclonal antibodies directed towards epitopes of one antigen, monoclonal antibodies directed towards epitopes of different antigens, polyclonal antibodies directed towards the same antigen, or polyclonal antibodies directed towards different antigens. Protocols may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Typically, an immunoassay for an anti-PilC antibody(s) will involve selecting and preparing the test sample suspected of containing the antibodies, such as a biological sample, then incubating it with an immunoreactive (also called antigenic) polypeptide(s) containing at least one epitope encoded in *pilC*. The incubation is under conditions that allow antigen-antibody complexes to form. Suitable incubation conditions are well known in the art. Subsequent to the incubation, complexes which are formed which contain the immunoreactive polypeptide are detected. The immunoassay

may be, without limitations, in a heterogenous or in a homogeneous format, and of a standard or competitive type.

In a heterogeneous format, the polypeptide is typically bound to a solid support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e.g., in membrane or microtiter well form), polyvinyl chloride (e.g., in sheets or microtiter wells), polystyrene latex (e.g., in beads or microtiter plates, polyvinylidine fluoride (known as Immulon), diazotized paper, nylon membranes, activated beads, and Protein A beads. For example, Dynatech Immulon[1] or Immulon[2] microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic polypeptide is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

In a homogeneous format, the test sample is incubated with antigen in solution. For example, it may be under conditions that will precipitate any antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of anti-PilC antibodies forming the antibody-antigen complex is directly monitored. This may be accomplished by determining whether labeled anti-xenogeneic (e.g., anti-human) antibodies which recognize an epitope on anti-PilC antibodies will bind due to complex formation. In a competitive format, the amount of anti-PilC antibodies in the sample is deduced by monitoring the

competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-PilC antibody (or, in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled anti-PilC antibodies in the complex may be detected using a conjugate of antixenogeneic Ig complexed with a label, (e.g., an enzyme label).

In immunoassays where PilC polypeptides are the analyte, the test sample, which may be a biological sample, is incubated with anti-PilC antibodies under conditions that allow the formation of antigen-antibody complexes. Various formats can be employed. For example, a "sandwich assay" may be employed, where antibody bound to a solid support is incubated with the test sample; washed; incubated with a second, labeled antibody to the analyte, and the support is washed again. Analyte is detected by determining if the second antibody is bound to the support. In a competitive format, which can be either heterogeneous or homogeneous, a test sample is usually incubated with antibody and a labeled, competing antigen is also incubated, either sequentially or simultaneously. These and other formats are well known in the art.

The antigenic regions of the polypeptides encoded in *pilC* can be mapped and identified by screening the antigenicity of expression products of *pilC* DNAs which encode portions of the PilC. The expression products may be from a variety of expression systems, including, for example bacterial systems, yeast systems, insect systems, and eukaryotic cell systems. In addition, studies giving rise to an antigenicity index

and hydrophobicity/hydrophilicity profile give rise to information concerning the probability of a region's antigenicity.

Efficient detection systems for infection with pathogenic microorganisms, (for example, *Neisseria, Pseudomonas, Bacteroides, Moraxella, or Vibrio*) may include the use of panels of epitopes. The epitopes in the panel may be constructed into one or multiple polypeptides. At least one of the epitopes will be encoded in *pilC* or derived from PilC. The assays for the varying epitopes may be sequential or simultaneous.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the polypeptides of the invention containing PilC epitopes (i.e., epitopes encoded within *pilC*) or antibodies directed against PilC epitopes in suitable containers. The kit may also contain other reagents, for example, buffer and standard, as well as other materials required for the conduct of the assay, as well as a suitable set of instructions for conducting the assay using the kit materials.

Another embodiment of the invention are oligomers. Using the disclosed portions of the *pilC* DNAs as a basis, oligomers of approximately 8 nucleotides or more can be prepared, either by excision from recombinant polynucleotides or by synthetic methods which are known in the art. These oligomers can serve as probes for the detection (including isolation and/or labeling) of polynucleotides which contain *pilC* sequences, and/or as primers for the transcription and/or replication of targeted *pilC* sequences. The oligomers contain a targeting polynucleotide sequence, which is comprised of nucleotides which are complementary to a target *pilC*

nucleotide sequence; the sequence is of sufficient length and complementarity with the *pilC* sequence to form a duplex which has sufficient stability for the purpose intended. For example, if the purpose is the isolation, via immobilization, of an analyte containing a target *pilC* sequence, the oligomers would contain a polynucleotide region which is of sufficient length and complementarity to the targeted *pilC* sequence to afford sufficient duplex stability to immobilize the analyte on a solid surface, via its binding to the oligomers, under the isolation conditions. For example, also, if the oligomers are to serve as primers for the transcription and/or replication of target *pilC* sequences in an analyte polynucleotide, the oligomers would contain a polynucleotide region of sufficient length and complementarity to a region flanking the targeted *pilC* sequence to allow the polymerizing agent to continue replication from the primers which are in stable duplex form with the target sequence, under the polymerizing conditions. The oligomers may contain a minimum of about 4 contiguous nucleotides which are complementary to a targeted *pilC* sequence; usually the oligomers will contain a minimum of about 8 contiguous nucleotides which are complementary to the targeted *pilC* sequence, and preferably will contain a minimum of about 14 contiguous nucleotides which are complementary to the targeted *pilC* sequence.

The oligomer, however, need not consist only of the sequence which is complementary to the targeted *pilC* sequence. It may contain in addition, nucleotide sequences or other moieties which are suitable for the purposes for which the oligomers are used. For example, if the oligomers are used as primers for the amplification of targeted *pilC* sequences via the

polymerase chain reaction (PCR), they may contain sequences which, when in duplex, form restriction enzyme sites which facilitate the cloning of the amplified sequences. Other types of moieties or sequences which are useful of which the oligomers may be comprised or coupled to, are those which are known in the art to be suitable for a variety of purposes, including the labeling of nucleotide probes.

In the basic nucleic acid hybridization assay, single-stranded analyte nucleic acid (either DNA or RNA) is hybridized to a nucleic acid probe, and resulting duplexes are detected. The probes for *pilC* sequences (natural or derived) are a length which allows the detection of these sequences by hybridization. While 6-8 nucleotides may be a workable length, sequences of 10-12 nucleotides are preferred, and about 20 nucleotides or more appears optimal. Preferably, these sequences will derive from regions which lack heterogeneity. These probes can be prepared using routine methods, including automated oligonucleotide synthetic methods. For use as probes, complete complementarity is desirable, although it may be unnecessary as the length of the fragment is increased.

For use of such probes as agents to detect the presence of *pilC* sequences, the sample to be analyzed (which may be biological) may be treated, if desired, to extract the nucleic acids contained therein. The resulting nucleic acid from the sample may be subjected to gel electrophoresis or other size separation techniques; alternatively, the nucleic acid sample may be dot blotted without size separation. In order to form hybrid duplexes with the targeting sequence of the probe, the targeted region of the analyte nucleic acid must be in single-stranded form. The latter may occur naturally;

alternatively, it may be accomplished by denaturation. Denaturation can be accomplished by various techniques known in the art. Subsequent to denaturation, the analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte, and the resulting duplexes containing probe(s) are detected.

Detection of the resulting duplex, if any, is usually accomplished by the use of labeled probes; alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., doxetanes, particularly triggered dioxetands), enzymes, antibodies, and the like. Variations of this basic scheme are known in the art.

If the targeted *pilC* sequences are expected to be present at relatively low levels, amplification may be required for their detection. Such techniques are known in the art. For example, the Enzo Biochemical Corporation "Bio-Bridge" system uses terminal deoxynucleotide transferase to add unmodified 3'-poly-dT-tails to a DNA probe. The poly dT-tailed probe is hybridized to the target nucleotide sequence, and then to a biotin-modified poly-A. PCT application 84/03520 and EPA124221 describe a DNA hybridization assay in which: (1) analyte is annealed to a single-stranded DNA probe that is complementary to an enzyme-labeled oligonucleotide; and (2) the resulting tailed duplex is hybridized to an enzyme-labeled

oligonucleotide. EPA 204510 describes a DNA hybridization assay in which analyte DNA is contacted with a probe that has a tail, such as a poly-dT tail, an amplifier strand that has a sequence that hybridizes to the tail of the probe, such as a poly-A sequence, and which is capable of binding a plurality of labeled strands. A particularly desirable technique may first involve amplification of the target *pilC* sequences. The target *pilC* sequences in sera may be amplified, for example, to approximately 10^6 sequences/ml. This may be accomplished, for example, by the polymerase chain reactions (PCR) technique described by Saiki et al. (1986), by Mullis, U.S. Patent No. 4,683,195, and by Mullis et al. U.S. Patent No. 4,683,202. Amplification may be prior to, or preferably subsequent to purification of the *pilC* target sequence. For example, amplification may be utilized in conjunction with the assay methods described in U.S. Patent No. 4,868,105, or if even further amplification is desired, in conjunction with the hybridization system in EPO Publication No. 317,077.

Generally, in the PCR technique, short oligonucleotide primers are prepared which match opposite ends of a desired sequence. The sequence between the primers need not be known. A sample of polynucleotide is extracted and denatured, preferably by heat, and hybridized with oligomers which are oligonucleotide primers, which are present in molar excess. Polymerization is catalyzed by a template- and primer-dependent polymerase in the presence of deoxynucleotide triphosphates (dNTPs), and may also be in the presence of nucleotide analogs. This results in two "long products" which contain the respective primers at their 5'-termini, covalently linked to the newly synthesized complements of the original strands. The replicated DNA is again

denatured, hybridized with oligonucleotide primers, returned to polymerizing conditions, and a second cycle of replication is initiated. The second cycle provides the two original strands, the two long products from cycle 1, and two "short products" replicated from the long products. The short products contain sequences (sense or antisense) derived from the target sequence, flanked at the 5'- and 3'-termini with primer sequences. On each additional cycle, the number of short products is replicated exponentially. Thus, this process causes amplification of a specific target sequence.

It will be understood that "primer", as used herein, may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the target region to be amplified. Hence, a "primer" includes a collection of primer oligonucleotides containing sequences representing the possible variations in the sequence or includes nucleotides which allow a typical base pairing. One of the primer oligomers in this collection will be homologous with the end of the target sequence.

The amplified sequence(s) may then be detected using a hybridization assay which utilize nucleic acid multimers which bind to single-stranded analyte nucleic acid, and which also bind to a multiplicity of single-stranded labeled oligonucleotides. A suitable solution phase sandwich assay which may be used with labeled polynucleotide probes, and the methods for the preparation of probes is described in EPO 225,807, published June 16, 1987.

The probes can be packaged into diagnostic kits. Diagnostic kits include the probe DNA, which may be labeled; alternatively, the probe DNA may be unlabeled

and the ingredients for labeling may be included in the kit in separate containers. The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, for example, standards, buffers, as well as instructions for conducting the test using the kit ingredients.

The *pilC* DNA sequence information in the clones described in the Examples may be used to gain further information on the remaining sequence of the *pilC* gene from meningococci, for other possible alleles of *pilC* in *Neisseria*, as well as *pilC* in other relevant genuses and species. This information will aid in the characterization of the gene, and of its role in virulence of the pathogenic forms of microorganisms, including, for example, *Neisseria*, *Pseudomonas*, *Bacteroides*, *Moraxella*, and *Vibrio*. Moreover, this sequence information can lead to additional polynucleotide probes, polypeptides derived from *pilC*, multiple *pilC loci*, and antibodies directed against PilC epitopes which would be useful for the diagnosis and/or treatment of infections caused by the relevant pathogenic microorganisms.

The DNA sequence information in the above-mentioned clones is useful for the design of probes for the isolation of additional DNA sequences which are derived from as yet undefined regions of *pilC*. For example, labeled probes containing a sequence of approximately 8 or more nucleotides, and preferably 20 or more nucleotides, which are derived from regions close to the ends of the DNA sequences shown in the Examples. These probes may be used to isolate overlapping DNA sequences within or adjacent to *pilC* from DNA libraries created from genomes of species having type 4 pilins. The resulting overlapping DNAs may then be used to

synthesize probes for identification of other overlapping fragments which do not necessarily overlap the DNAs whose sequences are given in the Examples. Thus, it is possible to sequence entire *pilC* genes utilizing the DNA sequences provided herein and the technique of isolation of overlapping DNAs derived from the *pilC* genes.

Methods for constructing DNA libraries are known in the art, and are discussed infra; for example, a method for the construction of *pilC* libraries in lambda-gt11 is discussed infra in Section IV.A. However, DNA libraries which are useful for screening with nucleic acid probes may also be constructed in other vectors known in the art, for example, lambda-gt10 (Huynh et al. (1985)). Another suitable vector for the creation of libraries may be EMBL3, which is a replacement vector which accepts inserts ranging from 9 to 23 kb in size. In general, methods for constructing DNA libraries is discussed in Maniatis et al, MOLECULAR CLONING, 2nd edition, (1989).

The sequence information derived from these overlapping *pilC* DNAs is useful for determining areas of homology and heterogeneity within the *pilC* gene(s), which could indicate the presence of different strains gonococci, meningococci, or other hitherto unrecognized pathogenic forms of *Neisseria*. It is also useful for the design of hybridization probes to detect PilC antigens or *pilC* nucleic acids in biological samples. Moreover, the overlapping DNAs may be used to create expression vectors for polypeptides derived from *pilC* gene(s).

The *pilC* DNA sequence information may also allow the construction of additional bacteriostatic agents for treatment of neisserial infections, in that they may block the expression of PilC and/or pilin assembly. For example, it may be used to derive

antisense polynucleotides. Antisense polynucleotides molecules are comprised of a complementary nucleotide sequence which allows them to hybridize specifically to designated regions of genomes or RNAs. Antisense polynucleotides may include, for example, molecules that will block protein translation by binding to mRNA, or may be molecules which prevent replication of DNA by replicase. They may also include molecules which carry agents (non-covalently attached or covalently bound) which cause the mRNA or genomic DNA to be inactive by causing, for example, scissions in these molecules. Antisense molecules which are to hybridize to *pilC* derived polynucleotides may be designed based upon the sequence information of the *pilC* DNA sequences provided herein, including those which would be isolated from additional DNA libraries. The antibacterial agents based upon anti-sense polynucleotides for *pilC* may be designed to bind with high specificity, to be of increased solubility, to be stable, and to have low toxicity. Hence, they may be delivered in specialized systems, for example, liposomes, or by gene therapy. In addition, they may include analogs, attached proteins, substituted or altered bonding between bases, etc.

Other types of drugs may be based upon polynucleotides which "mimic" important control regions of the *pilC* gene, and which may be therapeutic due to their interactions with key components of the system responsible for expression of the gene.

In addition to the specific methods described in the Examples, general methods are known which may be used in the practice of the invention. For example, general techniques used in extracting the genome from bacteria, including *Neisseria*, preparing and probing a DNA library, sequencing clones, constructing expression

vectors, transforming cells, performing immunological assays such as radioimmunoassays and ELISA assays, for growing cells in culture, and the like are known in the art and laboratory manuals are available describing these techniques. However, as a general guide, the following sets forth some sources currently available for such procedures, and for materials useful in carrying them out.

Both prokaryotic and eukaryotic host cells may be used for expression of desired coding sequences when appropriate control sequences which are compatible with the designated host are used. Among prokaryotic hosts, *E. coli* is most frequently used. Expression control sequences for prokaryotes include promoters, optionally containing operator portions, and ribosome binding sites. Transfer vectors compatible with prokaryotic hosts are commonly derived from, for example, pBR322, a plasmid containing operons conferring ampicillin and tetracycline resistance, and the various pUC vectors, which also contain sequences conferring antibiotic resistance markers. These markers may be used to obtain successful transformants by selection. Commonly used prokaryotic control sequences include the Beta-lactamase (penicillinase) and lactose promoter systems (Chang et al. (1977)), the tryptophan (trp) promoter system (Goeddel et al. (1980)) and the lambda-derived P_L promoter and N gene ribosome binding site (Shimatake et al. (1981)) and the hybrid tac promoter (De Boer et al. (1983)) derived from sequences of the trp and lac UV5 promoters. The foregoing systems are particularly compatible with *E. coli*; if desired, other prokaryotic hosts such as strains of Bacillus or Pseudomonas may be used, with corresponding control sequences.

Eukaryotic hosts include yeast and mammalian cells in culture systems. *Saccharomyces cerevisiae* and *Saccharomyces carlsbergensis* are the most commonly used yeast hosts, and are convenient fungal hosts. Yeast compatible vectors carry markers which permit selection of successful transformants by conferring prototrophy to auxotrophic mutants or resistance to heavy metals on wild-type strains. Yeast compatible vectors may employ the 2 micron origin of replication (Broach et al. (1983)), the combination of CEN3 and ARS1 or other means for assuring replication, such as sequences which will result in incorporation of an appropriate fragment into the host cell genome. Control sequences for yeast vectors are known in the art and include promoters for the synthesis of glycolytic enzymes (Hess et al. (1968); Holland et al. (1978)), including the promoter for 3 phosphoglycerate kinase (Hitzeman (1980)). Terminators may also be included, such as those derived from the enolase gene (Holland (1981)). Particularly useful control systems are those which comprise the glyceraldehyde-3 phosphate dehydrogenase (GAPDH) promoter or alcohol dehydrogenase (ADH) regulatable promoter, terminators also derived from GAPDH, and if secretion is desired, leader sequence from yeast alpha factor. In addition, the transcriptional regulatory region and the transcriptional initiation region which are operably linked may be such that they are not naturally associated in the wild-type organism. These systems are described in detail in EPO 120,551, published October 3, 1984; EPO 116,201, published August 22, 1984; and EPO 164,556, published December 18, 1985.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type

Culture Collection (ATCC), including HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells, and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include viral promoters such as that from Simian Virus 40 (SV40) (Fiers (1978)), Rous sarcoma virus (RSV), adenovirus (ADV), and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly A addition sequences; enhancer sequences which increase expression may also be included, and sequences which cause amplification of the gene may also be desirable. These sequences are known in the art.

Vectors suitable for replication in mammalian cells are known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding PilC epitopes into the host genome.

A vector which is used to express foreign DNA, and which may be used in vaccine preparation is Vaccinia virus. In this case the heterologous DNA is inserted into the Vaccinia genome. Techniques for the insertion of foreign DNA into the vaccinia virus genome are known in the art, and utilize, for example, homologous recombination. The insertion of the heterologous DNA is generally into a gene which is non-essential in nature, for example, the thymidine kinase gene (tk), which also provides a selectable marker. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (see, for example, Mackett et al. (1984), Chakrabarti et al. (1985); Moss (1987)). Expression of the polypeptide containing at least one immunoreactive PilC epitope then occurs in cells or individuals which are immunized with the live recombinant vaccinia virus.

Other systems for expression of desired polypeptides include insect cells and vectors suitable for use in these cells. These systems are known in the art, and include, for example, insect expression transfer vectors derived from the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcNPV), which is a helper-independent, viral expression vector. Expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive expression of heterologous genes. Currently the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed for improved expression. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; See Luckow and Summers (1989)).

Methods for the introduction of heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summer and Smith, Texas Agricultural Experiment Station Bulletin No. 1555; Ju et al. (1987); Smith et al. (1983); and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. The inserted sequences may be those which encode all or varying segments of the polyprotein, or other orfs which encode viral polypeptides. For example, the insert could encode the following numbers of amino acid segments from the polyprotein: amino acids 1-1078; amino acids 332-662; amino acids 406-662; amino acids 156-328, and amino acids 199-328.

The signals for posttranslational modifications, such as signal peptide cleavage, proteolytic cleavage, and phosphorylation, appear to be recognized by insect cells. The signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells. Examples of the signal sequences from vertebrate cells which are effective in invertebrate cells are known in the art, for example, the human interleukin 2 signal ($IL2_s$) which is a signal for transport out of the cell, is recognized and properly removed in insect cells.

Recombinant polynucleotides are inserted into host cells by transformation. Transformation may be by any known method for introducing polynucleotides into a host cell, including, for example packaging the polynucleotide in a virus and transducing a host cell with the virus, and by direct uptake of the polynucleotide. The transformation procedure used depends upon the host to be transformed. Bacterial transformation by direct uptake generally employs treatment with calcium or rubidium chloride (Cohen (1972); Maniatis (1989)). Yeast transformation by direct uptake may be carried out using the method of Hinnen et al. (1978). Mammalian transformations by direct uptake may be conducted using the calcium phosphate precipitation method of Graham and Van der Eb (1978), or the various known modifications thereof. Other methods for the introduction of recombinant polynucleotides into cells, particularly into mammalian cells, which are known in the art include dextran-mediated transfection, calcium phosphate mediated transfection, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the polynucleotides into nuclei.

The recombinant polynucleotide may be in the form of a vector. Vector construction employs techniques which are known in the art. Site-specific DNA cleavage is performed by treating with suitable restriction enzymes under conditions which generally are specified by the manufacturer of these commercially available enzymes. In general, about 1 microgram of plasmid or DNA sequence is cleaved by 1 unit of enzyme in about 20 microliters buffer solution by incubation of 1-2 hr at 37° C. After incubation with the restriction enzyme, protein is removed by phenol/chloroform extraction and the DNA recovered by precipitation with ethanol. The cleaved fragments may be separated using polyacrylamide or agarose gel electrophoresis techniques, according to the general procedures found in Methods in Enzymology (1980) 65:499-560.

Sticky ended cleavage fragments may be blunt ended using *E. coli* DNA polymerase I (Klenow) in the presence of the appropriate deoxynucleotide triphosphates (dNTPs) present in the mixture. Treatment with S1 nuclease may also be used, resulting in the hydrolysis of any single stranded DNA portions.

Ligations are carried out using standard buffer and temperature conditions using T4 DNA ligase and ATP; sticky end ligations require less ATP and less ligase than blunt end ligations. When vector fragments are used as part of a ligation mixture, the vector fragment is often treated with bacterial alkaline phosphatase (BAP) or calf intestinal alkaline phosphatase to remove the 5'-phosphate and thus prevent religation of the vector; alternatively, restriction enzyme digestion of unwanted fragments can be used to prevent ligation.

Ligation mixtures are transformed into suitable cloning hosts, such as *E. coli*, and successful

transformants selected by, for example, antibiotic resistance, and screened for the correct construction.

The desired recombinant DNA sequences may be synthesized by synthetic methods. Synthetic oligonucleotides may be prepared using an automated oligonucleotide synthesizer as described by Warner (1984). If desired the synthetic strands may be labeled with ^{32}P by treatment with polynucleotide kinase in the presence of ^{32}P-ATP, using standard conditions for the reaction.

DNA sequences, including those isolated from DNA libraries, may be modified by known techniques, including, for example site directed mutagenesis, as described by Zoller (1982). Briefly, the DNA to be modified is packaged into phage as a single stranded sequence, and converted to a double stranded DNA with DNA polymerase using, as a primer, a synthetic oligonucleotide complementary to the portion of the DNA to be modified, and having the desired modification included in its own sequence. The resulting double stranded DNA is transformed into a phage supporting host bacterium. Cultures of the transformed bacteria, which contain replications of each strand of the phage, are plated in agar to obtain plaques. Theoretically, 50% of the new plaques contain phage having the mutated sequence, and the remaining 50% have the original sequence. Replicates of the plaques are hybridized to labeled synthetic probe at temperatures and conditions which permit hybridization with the correct strand, but not with the unmodified sequence. The sequences which have been identified by hybridization are recovered and cloned.

DNA libraries may be probed using the procedure of Grunstein and Hogness (1975). Briefly, in this

procedure, the DNA to be probed is immobilized on nitrocellulose filters, denatured, and prehybridized with a buffer containing 0-50% formamide, 0.75 M NaCl, 75 mM Na citrate, 0.02% (wt/v) each of bovine serum albumin, polyvinyl pyrollidone, and Ficoll, 50 mM Na Phosphate (pH 6.5), 0.1% SDS, and 100 micrograms/ml carrier denatured DNA. The percentage of formamide in the buffer, as well as the time and temperature conditions of the prehybridization and subsequent hybridization steps depends on the stringency required. Oligomeric probes which require lower stringency conditions are generally used with low percentages of formamide, lower temperatures, and longer hybridization times. Probes containing more than 30 or 40 nucleotides such as those derived from cDNA or genomic sequences generally employ higher temperatures, e.g., about 40-42°C, and a high percentage, e.g., 50%, formamide. Following prehybridization, 5'-^{32}P-labeled oligonucleotide probe is added to the buffer, and the filters are incubated in this mixture under hybridization conditions. After washing, the treated filters are subjected to autoradiography to show the location of the hybridized probe; DNA in corresponding locations on the original agar plates is used as the source of the desired DNA.

For routine vector constructions, ligation mixtures are transformed into *E. coli* strain HB101 or other suitable host, and successful transformants selected by antibiotic resistance or other markers. Plasmids from the transformants are then prepared according to the method of Clewell et al. (1969), usually following chloramphenicol amplification (Clewell (1972)). The DNA is isolated and analyzed, usually by restriction enzyme analysis and/or sequencing. Sequencing may be by the dideoxy method of Sanger et al. (1977) as further

described by Messing et al. (1981), or by the method of Maxam et al. (1980). Problems with band compression, which are sometimes observed in GC rich regions, were overcome by use of T-deazoguanosine according to Barr et al. (1986).

An enzyme-linked immunosorbent assay (ELISA) can be used to measure either antigen or antibody concentrations. This method depends upon conjugation of an enzyme to either an antigen or an antibody, and uses the bound enzyme activity as a quantitative label. To measure antibody, the known antigen is fixed to a solid phase (e.g., a microplate or plastic cup), incubated with test serum dilutions, washed, incubated with anti-immunoglobulin labeled with an enzyme, and washed again. Enzymes suitable for labeling are known in the art, and include, for example, horseradish peroxidase. Enzyme activity bound to the solid phase is measured by adding the specific substrate, and determining product formation or substrate utilization colorimetrically. The enzyme activity bound is a direct function of the amount of antibody bound.

To measure antigen, a known specific antibody is fixed to the solid phase, the test material containing antigen is added, after an incubation the solid phase is washed, and a second enzyme-labeled antibody is added. After washing, substrate is added, and enzyme activity is estimated colorimetrically, and related to antigen concentration.

Examples

Described below are examples of the present invention which are provided only for illustrative purposes, and not to limit the scope of the present invention. In light of the present disclosure, numerous

embodiments within the scope of the claims will be apparent to those of ordinary skill in the art.

Example 1

Isolation of PilC

An outer membrane preparation from *N. gonorrhoeae* strain $MS11_{ms}$ (P^+) contains small amounts of a 110 kd protein, PilC. This protein was enriched during alternate cycles of crystallization and solubilization of pili, unlike other outer membrane proteins that decreased in abundance by this procedure.

The materials and methods used for the isolation procedure were the following.

Bacterial strains and growth conditions

N. gonorrhoeae $MS11_{ma}$ (Meyer et al., 1984) and P^- and P^-n variants of $MS11_{mk}$ (Swanson et al., 1986) were kindly obtained from Dr. M. So and from Dr. M. Koorney, respectively. The gonococcal isolates UM01 and KH4318 have previously been described (Norlander et al., 1981). *N. gonorrhoeae* strains 605344 and 605103 were obtained from Dr. D. Danielsson, Örebro, Sweden, and strain 765 was isolated at the Department of Bacteriology in Umeå, Sweden. The commensal *Neisseria* species *N. lactamica* Nctc 10618 and *N. subflava* GN01 were obtained from Pharmacia, Uppsala, Sweden. These bacteria were grown at 37°C in a 5% CO_2 atmosphere on Difco GCB agar containing Kellogg's supplement. Piliated (P^+) and nonpiliated (P^-) variants were distinguished by colony morphology and passed as single colonies. *E. coli* strain Y 1090 (obtained from Promega Biotech) was used for plaque screening, DH5 (Hanahan, 1985) for molecular cloning, AA10 *recA* (Stoker et al. 1984) for isolation of minicells and TG1 (Gill et al., 1986) for propagation of M13 clones.

Preparation of pili and outer membranes

Pili were prepared essentially as described by Brinton et al. (1978). Gonococci (P^+Tr) from 80 GGB plates, grown for 18 h, were harvested in 0.05 M Tris-HCl pH 8.0 and 0.15 M NaCl, washed twice and resuspended in 40 ml 0.15 M ethanolamine pH 10.5. Pili were sheared off in a Sorvall Omnimixer, setting 3 for 30 s. The cell debris was pelleted at 13,000 g for 30 min at 4°C and the supernatant was dialyzed against 0.05 M Tris-HCl pH 8.0 and 0.15 M NaCl. The crystallized pili were pelleted at 13,000 g for 60 min., resuspended in 0.15 M ethanolamine pH 10.5, and centrifuged at 23,000 g for 60 min. The supernatant was dialyzed as described above against 0.05 M Tris-HCl pH 8.0 and 0.15 M NaCl. Several cycles of crystallization and solubilization were performed to produce pili preparations with high purity. Outer membranes of *N. gonorrhoeae* were prepared by the sarkosyl method described by Norquist et al. (1978).

Example 2

Preparation of Purified Anti-PilC Antibodies

The 110 kd protein present in purified $MS11_{ms}$ pili preparations was eluted from SDS polyacrylamide gels and rabbit antibodies were generated against the gel purified protein. The antiserum cross reacted extensively with the pilin protein in immunoblots and was therefore absorbed with extracts of *Pseudomonas putida* expressing the pilin subunit of *N. gonorrhoeae* on plasmid pGC02.

Pili preparations of *N. gonorrhoeae* $MS11_{ma}$ (P^+) crystallized 5 times were separated on 10% SDS-polyacrylamide gels using the buffer system of Laemmli (1970). These gels were stained in 0.25 M KCl and 1 mM

DTT for 5 min., the 110 kd protein band was sliced out, crushed and incubated in a buffer containing 0.05 M Tris-HCl pH 7.9, 0.1 mM EDTA, 5 mM DTT and 0.15 M NaCl at 4°C overnight. Gel pieces were removed by centrifugation prior to immunization of rabbits.

The achieved 110 kd-antiserum was extensively absorbed with *Pseudomonas putida* 2440 (Bagdasarien et al. (1983), carrying a recombinant plasmid, pGCO2, constructed as follows. The 1.0 kb *Hpa*I-*Eco*RI fragment of the pilus gene clone pNG1100 (Meyer et al. 1984) obtained from M. So was cloned into the *Hpa*I and *Eco*RI sites of pMMB66 (Fürste et al., 1986). The *pilE* gene is then under control of the *tac* promoter and induction with 1 mM IPTO resulted in high levels of pilin produced in *P. putida* 2440, but no extracellular pili structures were observed. Dense sonicated cultures of *P. putida* 2440/pGC02 were mixed in a 1:1 ratio with the crude antiserum. About 15 cycles of 1 h incubation and 30 min centrifugation at 25,000 g in the presence of 1 mM PMSF (phenylmethylsulfonylfluoride) at 4°C were performed.

The pili antiserum used in immunoblots was generated in a rabbit against highly purified pili preparations of *N. gonorrhoeae* $MS11_{ma}$.

In immunoblots 10 µg of boiled bacterial cells or the same amount of outer membranes were electrophoresed on 10% SDS-polyacrylamide gels. The proteins were transferred from the gel onto nitrocellulose sheets where their immunological cross-reaction with the 110 kd absorbed antiserum was tested using an immunoblotting protocol as described by Towbin et al. (1979).

Example 3

Specificity of Purified Anti-PilC Antibodies

The absorbed antiserum was used in immunoblots with whole cell extracts of a number of *N. gonorrhoeae* strains as well as commensal strains of Neisseria (Figure 2A). All strains of *N. gonorrhoeae*, except strain 605103, contained one or two high molecular weight protein species reacting with the antiserum. Strain 605103, unlike the other strains tested, was nonpiliated and no piliated variants could be obtained suggesting that it is a P^-n variant (Swanson et al., 1985). This was confirmed by Southern blot hybridization using an oligonucleotide probe corresponding to the 5' end of the *pilE* gene. No hybridization was obtained with this probe. The commensal *N. lactamica* Nctc10618, but not *N. subflava* GN01, contained a high molecular weight protein reacting with the 110 kd antiserum. Immunoblots against outer membrane preparations of P^+ and P^-n $MS11_{mk}$ showed the 110 kd protein to be present in the outer membrane in both of these MS11 variants.

Southern blot hybridization was accomplished as follows. Digested genomic DNA was separated on 0.7% agarose gels and transferred to nitrocellulose filters (Southern, 1975). After transfer and baking the filters were prehybridized in a mixture of 5 x SSC, 0.1% SDS, 5 mM EDTA, 5 x Denhardt's solution and 100 μg/ml of sonicated calf thymus DNA at 65°C for 2-6 h. ^{32}P-labeled probe (multiprime DNA labelling system, Amersham International) was added and hybridization was performed for 12-15 h at the same temperature. The filters were washed in 2 x SSC with 0.1% SDS and in 0.2 x SSC with 0.1% SDS for 2 x 15 min each, dried and exposed to Kodak XRP film at -80°C.

A 21-base-long oligonucleotide complementary to the signal peptide coding region of *pilE* (5'-GCCTTTTTGAAGGGTATTCAT-3') was ^{32}P-labeled with T4 polynucleotide kinase and used to probe *Cla*I-digested genomic DNA. The blot was prehybridized at 37°C in a mixture containing 2 x Denhardt's, 0.1% SDS, 2.5 mM EDTA, 5 x SSC and 100 µg/ml sonicated calf thymus DNA, hybridized at 37°C and washed in 2 x SSC for 5 min. $MS11_{mk}(P^+)$ gave a 4 kb hybridization fragment, whereas $MS11_{mk}(P^-n)$ and 605103 gave no hybridization signal.

Example 4

Molecular cloning of the *pilC1* gene encoding a 110 kd protein

Chromosomal DNA from *N. gonorrhoeae* $MS11_{mk}(P^+)$ was used to construct a λgt11 library. The library was screened with the absorbed 110 kd antiserum and one positive clone out of 10,000 plaques was found, containing an 800 bp insert. A lysogen of this positive λgt11 clone was examined in immunoblots and a fusion protein with an estimated size of 150 kd reacted with the antiserum (data not shown). The 800 bp insert was purified, labeled with ^{32}P, and used as a probe to screen a plasmid library from *N. gonorrhoeae* $MS11_{ms}$. Six clones out of 10,000 hybridized with the probe. Restriction maps for these partially overlapping six clones are shown in Figure 1.

In Figure 1, plasmids pABJ04-09, which all belong to locus 1, were isolated from a plasmid library using the 800 bp insert from λgt11 as a probe. The λgt11 insert (from locus 2) has an additional *Sal*I site not found in the plasmid clones. The position of the *pilC1* gene and direction of its transcription (indicated by an arrow) were determined in *E. coli* minicells. Three

thickened lines, with an arrow-head at each end, indicate fragments used as probes in Southern hybridizations. I.e., the 800 bp insert from λgt11, the *EcoRV*$_1$-*EcoRV*$_2$ (1.3kb) and the *EcoRV*$_3$-*HindIII*$_4$ (0.8 kb) fragments of pAGJ04. Triangles mark the location of two mTnCm insertions in pABJ04. The resulting plasmids, pABJ)4::mTnCm-12 and pABJ04::mTnCm-14 were used, were used to inactivate *pilC1* and *pilC2*.

The six plasmid clones, pABJ04-09, were transformed into the minicell producing strain AA10 to monitor expression of plasmid encoded [^{35}S]methionine labeled proteins. The *E. coli* minicell strain AA10 was transformed with plasmid DNA (pABJ04-09) and chromosome deficient minicells from these strains were purified over sucrose gradients (Thompson and Achtman, 1978). The plasmid-encoded proteins were labeled in the presence of 80 μCi [^{35}S]methionine in minimal salts medium and 1% methionine assay medium (Difco). After lysis of the minicells in sample buffer (Laemmli, 1970) the proteins were electrophoresed on an SDS-polyacrylamide gel, the gel was dried and exposed to X-ray film (Kodak X-OmatAR).

Plasmid pABJ04 expressed minute amounts of three high molecular weight proteins, 113, 111 and 108 kd in size, as well as a number of lower molecular weight protein species not produced from the vector control. The three high molecular weight bands were missing in pABJ05 and pABJ06 but three novel lower molecular weight protein species had appeared, suggesting that pABJ05 and pABJ06 are deleted for the 3' end of a gene, denoted *pilC1*, and that this gene is responsible for all three high molecular weight species. This suggested that the distal end of the gene must be located between the *MluI*$_1$ and *MulI*$_2$ sites (Figure 1). The observation that plasmid pABJ07 did not express any high molecular proteins

tentatively located the 5' end of the gene to a region 0.5-1.2 kb to the right of the *Eco*RV$_3$ site. The size for a gene encoding a 110 kd protein is ~3 kb which is in agreement with these mapping data.

Example 5

Identification of a Second Gene Encoding PilC

The 800 bp insert in λgt11 contains a single *Sal*I site not present in the region on pABJ04 which hybridized to this fragment, suggesting that there is more than one *pilC* locus in the genome of *N. gonorrhoeae* MS11. This was confirmed in Southern blot hybridizations in which three different *pilC* fragments were used to probe *Sma*I and *Cla*I digested genomic DNA. The 800 bp fragment from λgt11 hybridized in a Southern blot to two *Cla*I (18 and 8 kb) and *Sma*I (13 and 4.5 kb) fragments of DNA prepared from *N. gonorrhoeae* MS11$_{mk}$. Since the probe does not contain any internal *Cla*I or *Sma*I sites, there are presumably two copies of the 3' end of *pilC* in the MS11 genome. The 1.3 kb *Eco*RV$_1$ - *Eco*RV$_2$ fragment of pABJ04 carries the central region of *pilC1*. This probe hybridized to the same two *Cla*I fragments and to four *Sma*I fragments, two of which are the same size as the two *Sma*I fragments identified with the 800 bp probe (13 kb and 4.5 kb). Hybridization with the 800 bp probe was more extensive to the 8 kb *Cla*I and the 4 kb *Sma*I fragment whereas the reverse was found with the 1.3 kb *Eco*RV$_1$ - *Eco*RV$_2$ fragment from pABJ04 strongly suggesting that the two genomic copies of *pilC* show a significant sequence variation in the 3' as well as in the central region. A probe corresponding to the 5' region of *pilC1* was also used in Southern hybridization experiments. This 0.8 kb *Hin*dIII$_4$ - *Eco*RV$_3$ fragment hybridized to two *Cla*I (18 kb and 4 kb) and *Sma*I (25 kb and 7 kb) fragments

with seemingly equal efficiency. The hybridization pattern was identical using DNA from $MSII_{ms}$. Taken together these hybridization data indicate that *N. gonorrhoeae* MS11 contains two complete copies of *pilC*. Furthermore the two genes appear to be more homologous in their 5' as compared to their central and 3' regions.

The results indicate that the 800 bp insert from λgt11 carries information from *pilC2* whereas the clones pABJ04-09 must carry information from *pilC1*. Finally *pilC2* must be located >2 kb from either end of *pilC1*. The DNA sequence of the 3'-end of the *pilC2* fragment is shown in Figure 7. The sequence showing the putative amino acids encoded therein are shown in Figure 8. A comparison of the analogous portions of *pilC2* (top) and *pilC1* (bottom) DNA sequences, and the putative amino acids encoded therein are shown in Figure 9.

The 800 bp fragment from *pilC2* was also used to probe digested genomic DNA from *N. gonorrhoeae* strains UM01, 765 and 605103. The latter isolate does not express detectable levels of the 110 kd protein. Strain UM01, unlike MS11, contained only one *Cla*I fragment of 15 kb that hybridized to the probe (data not shown). Hence, this strain may contain only one copy of *pilC*. Strain 605103 and 765, on the other hand, each seem to contain two copies of *pilC* since two *Cla*I and two *Sma*I fragments hybridized to the 800 bp probe.

The commensal *N. lactamica* Nctc10618 DNA digested with *Cla*I and *Sma*I also hybridized with the 800 bp probe. Since only one band hybridized in each case this strain may contain only one copy of *pilC*. In contrast, *N. subflava* GN01 did not hybridize to the 800 bp *pilC2* probe using the same stringency.

Example 6
Characterization of the *pilC* Genes

The pilC1 gene on pABJ04 is translationally out of frame

The amino terminal sequence of gel purified 110 kd protein from strain $MS11_{ms}$ (P^+) was determined by sequential Edman degradation. For aminoterminal sequence determination automated Edman degradations (Edman and Bregg, 1967) were performed in an updated Beckman 890C spinning cup sequencing sequencer. The sequencing procedure and the method for analysis of the 3-phenyl-2-thiohydantoin derivatives been described (Engström et al., 1984). Considerable difficulties were encountered in the method probably due to blocking of the N-terminus. As a result, only the residues from position 4 to 10 were obtained (Figure 2).

The 3.3 kb *Hin*dIII$_4$ - *Mlu*I$_1$ fragment encompassing the entire *pilC1* gene was sequenced on both strands using the dideoxy sequencing method adapted for single stranded DNA.

Purified DNA fragments from pABJ04 and PCR-amplified 5' end of *pilC1* was subcloned into M13 vectors (Sanger et al., 1980: Yanish-Perroa et al., 1985) and sequenced using the chain termination method of Sanger et al. (1977). Primers used were the M13 17-mer universal primer and oligonucleotides synthesized at Symbicon, Umeå, Sweden or at the Department of Biochemistry, Washington University, St., Louis, MO, USA.

The results of the sequencing showed that the *pilC1* contained one single open reading frame of 997 codons (from left to right in Figure 2) and starting at an AUG codon 195 bp from the *Hin*dIII$_4$ site. Codons 7-12 in this open reading frame corresponded to amino acids 5-10 in the sequence of the gel purified protein. The AUG

codon in the beginning of the long open reading frame was not preceded by a typical Shine-Dalgarno sequence. Moreover, since the 110 kd protein is located in the outer membrane of *N. gonorrhoeae*, we expected the protein to be translated with a signal sequence. When examining the nucleotide sequence, an AUG codon was found in frame 1 that was preceded by a typical Shine-Dalgarno sequence (-AGGAA-). The sequence following this AUG codon would encode a typical signal peptide with basic amino acids in the amino terminal region and a hydrophobic central region. However, no signal peptidase cleavage site could be predicted following the rules of von Heijne (1983). A tract of 12G residues was found in the region encoding the putative signal peptide for PilC. Addition of one G residue or the loss of two would align the long open reading frame with the AUG codon in frame 1. The translated region in frame 2 contains a putative signal peptidase cleavage site between Ala and Gln. A cleavage at this site would align the determined amino acid sequence at positions 5-10 for the 110 kd protein with the deduced amino acid sequence. The data therefore suggested that the cloned *pilC1* gene is out of frame due to frameshifting in the region encoding the signal peptide.

Figure 2 shows the nucleotide sequence and the deduced amino acid sequence of the 5'-end of *pilC*. The amino-terminal sequence of gel purified PilC from $MS11_{ms}(P^+)$ is shown in a box below frame 2, a 997 amino acid long open reading frame that would code for a protein about 110 kd in size. Frame 1 contains 41 amino acids and is preceded by a putative Shine-Dalgarno sequence (underlined). Two horizontal lines mark a stretch of 12 G residues. An addition of one G in this region would align the ATG (boxed) in frame 1 with frame

2. Numbers above the sequence show base positions relative to the *Hin*dIII$_4$ site (=0) located on pABJ04. The position of two 24 bp oligonucleotide primers (opposite stands) used for PCR amplification, are indicated above the sequence by hatched bars.

Figure 3 shows the nucleotide sequence of the sense strand of the *pilC1* gene.

Figure 4 shows the shows the nucleotide sequence of the sense strand of the *pilC1* gene and the amino acids encoded therein.

Genetic inactivation of *pilC2* but not *pilC1* abolishes expression of the 110 kd protein in MS11

Plasmid pABJ04 was mutagenized in *E. coli* by a transposon mini-Tn3 derivative, mTnCm. The shuttle mutagenesis system developed by Seifert et al., (1986) using a miniTn3 carrying the chloramphenicol resistance gene was kindly provided by Dr. M. So. Mutagenesis of pABJ04 with mTnCm and transformation of *N. gonorrhoeae* were performed as previously described (Seifert et al., 1990). MiniTnCm insertions at 30 different positions in pABJ04 were identified, two of which mapped within the PilC gene. Piliated *N. gonorrhoeae* MS11$_{mk}$ were transformed with 2 μg plasmid DNA, transformants were selected for on plates containing 10 μg/ml chloramphenicol for the single mutants and 30 μg/ml chloramphenicol for the double mutants.

Only two mTnCm insertions had occurred in *pilC1* (Figure 1). Truncated protein species were seen in minicells with the mTnCm-14 insertion located 0.5 kb from the 3′ end of *pilC1* but not with the mTnCm-12 insertion located 0.5 kb from the 5′ end of the gene. Both insertion mutants were used in a gene replacement experiment. Plasmids pABJ04::mTnCm-12 and pABJ04::mTnCm-

14 were linearized with BamHI and transformed into *N. gonorrhoeae* $MS11_{mk}(P^+)$ and transformants resistant to 10 μg/ml of chloramphenicol were selected. Forty-eight P^+ transformants (24 from each experiment) were assayed for the presence of PilC in immunoblots. All these transformants remained capable of expressing the PilC protein. Genomic DNA was prepared from seven of the chloramphenicol resistant transformants (five from pABJ04::mTnCm-12 and two from pABJ04::mTnCm-14), cleaved with ClaI and PvuII and used in Southern blot experiments using the $EcoRV_1$ - $EcoRV_2$ fragment of pABJ04 as a probe. The 8 kb ClaI fragment was unaffected in the mutants whereas the 18 kb ClaI fragment had been replaced by a 20 kb fragment. PvuII cleaves within the 1.6 kb mTnCm element. The probe detected an 8 kb PvuII fragment in both parent and mutant DNA. In the mutants, a novel PvuII fragment appeared that was 6.2 kb in size in five transformants obtained with pABJ04::mTnCm-12 and 4.8 kb in size in two transformants with pABJ04::mTnCm-14. To confirm the insertion of mTnCm, a 250 bp EcoRI-HindIII fragment of the CAT GenBlock (Pharmacia, Sweden), containing the PvuII site, was used as a probe. It detected the larger of the two ClaI fragments as well as the 6.2 kb PvuII fragment. In addition, a 2 kb PvuII fragment not covered with the *pilC* probe was detected. These data demonstrate that we have obtained gene replacements in *pilC1*, whereas *pilC2* was unaffected in all seven P^+, $PilC^+$ transformants. A rapid hybridization was done to screen the remaining 41 P^+ transformants. All but one had mTnCm inserted in *pilC1*. The remaining transformant had an intact locus 1 and 2 and must therefore contain mTnCm elsewhere in the gonococcal chromosome.

In the same transformation experiments, the frequency of P^- colony variants was about five-fold higher as compared with that occurring normally in strain $MS11_{mk}(P^+)$. Two P^-mTnCm-12 transformants isolated at 10 μg/ml of chloramphenicol were also analyzed by Southern blot hybridization using the $EcoRV_1$ - $EcoRv_2$ fragment of pABJ04 and the EcoRI - HindIII fragment of the CAT GenBlock. Each of these mutants carried mTnCm in *pilC2* as evidenced by a replacement of the 8 kb ClaI fragment by a fragment 9.5 kb in size that hybridizes to both probes. These *pilC2*::mTnCm insertion mutants did not express PilC as determined by immunoblot analysis.

A P^+, *pilC1*::mTnCm-12 mutant was retransformed with DNA prepared from a P^-, *pilC2*::mTnCm-12 mutant and colonies growing at 30 μg/ml of chloramphenicol were selected to obtain double mutants in *pilC*. All resistant transformants were P^-, and when analyzed by Southern blot hybridization all contained mTnCm in both *pilC1* and *pilC2*. Electron microscopy revealed that the P^+, *pilC1*::mTnCm-12 mutant still expressed pili albeit at a slightly lower level than the $MS11_{mk}(P^+)$ parental clone, whereas the P^-, *pilC2*::mTnCm-12 was completely bald as was the *pilC1*, *pilC2* double mutant.

Immunoblot analyses were performed on the P^+ *pilC1*::mTnCm-12 mutant, the P^-, *pilC2*::mTnCm-12 mutant and the P^-, *pilC1*::mTnCm-12, *pilC2*::mTnCm-12 double mutant, using PilC and pili antisera. Inactivation of *pilC1* did not abolish expression of PilC or the pilin. Inactivation of *pilC2* totally abolished expression of PilC but did not affect expression of pilin. The *pilC1*, *pilC2* double mutant was $PilC^-$ but produced only low levels of pilin. Taken together these data imply that *pilC2* but not *pilC1* is expressing PilC in the MS11

variant under study. Moreover, inactivation of *pilC2* but not *pilC1* was associated with a loss of piliation.

P^+ revertants occurred spontaneously at a low frequency in the *pilC2*::mTnCm-12 mutants. These revertants expressed pili as determined by electron microscopy and also expressed PilC. It is likely that PilC expression is due to in-frame switching in *pilC1*.

Figure 5 shows the nucleotide sequence of the sense strand of the *pilC1* gene, and the effect of frame shift on the putative gene products encoded therein.

The *pilC* genes of *N. gonorrhoeae* vary in the length of the G tract

Polymerase chain reaction (PCR) with Taq polymerase was used to analyze the 5' region of *pilC* using two 24 base long synthetic oligonucleotides based on the sequence of *pilC1* (Figure 2). These oligonucleotides would generate an amplified fragment of 149 bases as judged from the sequence obtained from pABJ04.

Polymerase chain reaction was carried out in 100 µl containing 50 ng of genomic DNA or 5 ng of plasmid DNA. 1.0 µM of each oligonucleotide, 200 µM of each nucleotide, 0.001% gelatin, 1.5 mM $MgCl_2$, 10 mM Tris (pH 8.3), 50 mM KCl, 0.25 µl 1 mCi/ml [^{32}P]dATP and 2 U of Taq Polymerase (Perkin Elmer Cetus). The samples were passed through 25 cycles: 2 min at 50°C, 1 min at 94°C and 3 min at 72°C in a Thermal Cycler (Perkin Elmer Cetus). Aliquots of the DNA fragments were denatured at 95°C for 2 min and electrophoresed on standard denaturing sequencing gels.

The amplified products from $MS11_{mk}$ (P^+) DNA were 149 and 150 long respectively. In addition, two less abundant products of 151 and 148 bases were seen. The

amplified products were electroeluted and cloned into M13mp18, and twenty phage clones were sequenced using a universal primer. Four different sequences were obtained (Figure 6).

Figure 6 shows the nucleotide sequence of PCR amplified fragments demonstrating a variation in the length of the G tract and sequence differences in the 5'-region of the *pilC* genes. The two oligonucleotide primers used for the PCR are shown in Figure 2. Amplified DNA was cloned into M13mp8 and sequenced. Shown are the complete nucleotide sequence in between the two primers. In-frame sequences are translated and the G stretches are underlined. The putative cleavage sites are marked with arrows. Genomic DNA from *N. gonorrheae* strains MS11(P^+.Pilc$^+$), UM01(P^+.Pilc$^+$), 765 (P^+.Pilc$^+$) and 605103 (P^-n.Pilc$^-$), and purified DNA from pABJ04/AA10(*recA*) was used in the PCR.

Variant patterns 1a and 1b were identical to each other and to the cloned sequence on pABJ04 except for the presence of 11 instead of 12 G residues in the G tract of 1b. The G tract of sequence 2a was 13 residues long indicating that the sequence is in frame. In addition, this sequence differed from *pilC1* by four basepair substitutions outside the G tract, including an AAA lysine codon four triplets downstream of the putative signal peptide processing site which is in agreement with the lysine residue found in the fourth position of the gel purified 110 kd PilC protein. Sequences 1a and 1b contained CAA, the codon for Gln, at the same position. Sequence 2b was identical to 2a except for the presence of 12 G residues in the G tract. These data are compatible with sequence 1 being from *pilC1* and sequence 2 from *pilC2* and further support that *pilC2* must be the expressed locus in the MS11(P^+) variant we are studying.

Strain UM01 apparently only contains one copy of *pilC*. DNA from this strain generated five amplified fragments ranging in size from 148 to 152 bp in the PCR reaction. The most abundant fragments were 149-151 bp long. Among ten M13 clones, three variant 1 sequences were found (a,b,c) that differed only in the number of G residues (11-13) in the G tract (Figure 6) supporting the hybridization data that this strain contains only one *pilC* gene. Since a PilC protein is expressed from UM01 we suggest that the majority of cells has 13 Gs in the G tract.

Strain 765 contains two *pilC* loci, both of which seem to be translationally ON based on the presence of two high molecular weight proteins reacting with the absorbed PilC antiserum. A number of amplified fragments were seen after the PCR reaction ranging in size from 149 to 153 bases. Three variant sequences were found among nine clones (Figure 6). The G tract was 13 residues long in variant 3a (in frame) and 14 (out of frame) in variant 3b whereas sequence variant 4 contained 11 G residues in the G tract. Variant sequence 4 contained four additional nucleotides (-CAGG-) distal to the G tract relative to variant sequences 1, 2 and 3, indicating that the amplified product with 11 Gs from this variant sequence is 152 long and out of frame. Two PCR amplified products 152 and 153 in length were obtained from strain 765 suggesting that in frame variants of sequence 4 might be present in the DNA prepared from this strain.

Strain 605103 carries two *pilC* copies, both of which seem to be translationally OFF. The amplified fragments were 148 and 149 bases in size. Out of eight M13 clones only variant 1a and 1b sequences were found, with 11 and 12 Gs in the G tract respectively. Consequently, we were unable to find an in frame sequence

variant from this strain. We do not know if the 5' ends of the two *pilC* genes are identical in this strain or if one *pilC* gene differed from *pilC1* in the region corresponding to the oligonucleotides used for amplification. In the latter case we would not expect to obtain any amplified products from the second copy.

The only in frame variant found in DNA amplified from *N. gonorrhoeae* carried 13 Gs in the G tract. To see if variants with 10Gs arise in products expressed in *E. coli*, PCR amplified products were generated from pABJ04 purified from *E. coli* strain AA10, using the same two oligonucleotide primers as before. Out of 12 sequenced clones, two carried 10 Gs in the G tract (Figure 6). The majority of clones (seven) carried 12 Gs as expected. It is likely that the PCR amplification products are not representative of the original DNA population. However, the distribution of variation in the G tract is consistent with a model in which only one G residue is gained or lost at one given event. Since AA10 is *recA*, frameshift mutations in the G tract in *E. coli* occur independently of the RecA protein.

N. meningitidis contains two *pilC* loci

Southern blot hybridizations using MS11 *pilC1*⁻ specific probes identified multiple fragments when meningococcal genomic DNA is digested with a variety of restriction endonucleases. PCR amplification using two 24-base oligonucleotides from the 5' end of MS11 *pilC1* as primers yields multiple fragments ranging in size from 148 to 151 bases. DNA sequencing of fragments cloned into phage M13 identifies two classes of sequences, as in *N. gonorrhoeae*, which differ outside the G-tract. Variation occurred within each class with respect to the number of G's in the G-tract. Therefore, *N. meningitidis* must

carry two *pilC* loci, as does *N. gonorrhoeae*, which should also be under the control of translational frame shifting.

Both *pilC* loci are cloned from *N. meningitidis* by generating an EMBL3 library and screening this library with *pilC1*$^-$ and *pilC2*$^-$ specific DNA from *N. gonorrhoeae* strain MS11.

Genomic DNA from *N. meningitidis* is partially digested with Sau3A and fragments ranging from 9 to 20 kB are ligated into the lambda EMBL3 vector. Because of the packaging constraints of the phage, only those lambdas which contain DNA fragments of this size will be packaged (i.e., are viable). The library thus constructed can be screened with genomic oligonucleotide or cloned gene probes following selection in a lysogenic *E. coli* strain. (Frischart, A.M. et al (1983), J. Mol. Biol. 170:827). Preferably, full length clones are identified by screening for clones hybridizing to both the 5′ and 3′ ends of *pilC*. If full length clones cannot be obtained from the EMBL3 library, *pilC* specific probes may be used to screen a plasmid library from the same strains.

Translational fusion proteins with β-galactosidase may also be screened for in a λgt11 library, using β-galactosidase and PilC$^-$ specific antisera in Western immunoblots. β-galactosidase-PilC1 and β-galactosidase-PilC2 fusion proteins are purified from the cytoplasm of recombinant *E. coli* and used to raise specific antisera.

Example 7

The immunobiological properties of PilC

PilC is located in the outer membrane of *Neisseria*. The immune response during natural infection can be assessed by screening convalescent sera for anti-

PilC antibody. The presence of only two *pilC* loci suggests that PilC is only moderately variable, however. This together with its essential role in pilus biogenesis makes PilC attractive as a potential vaccine candidate.

Two types of PilC translational fusions using alkaline phosphatase and β-galactosidase are generated. In the first instance a secreted fusion protein is obtained that may associate with the outer membrane. In the second instance the fusion proteins may accumulate in the cytoplasm as inclusion bodies. The construction schemes for such fusion proteins uses techniques known in the art. Tn*phoA* insertions on plasmid pABJ04 in *E. coli* are generated, and a PhoA$^+$ phenotype is screened for as blue colonies on media containing the chromogenic substrate XP. If such clones have the *phoA* gene in frame with an in frame variant of *pilC1* the fusion product should be able to cross the cytoplasmic membrane where it can be analyzed by Western immunoblots using an alkaline phosphatase specific antiserum and our PilC antiserum raised against gel-purified PilC2 from MS11(P$^+$). *LacZ::pilC* fusions are generated by cloning different segments of *pilC* into a *lacZ*$^-$ containing vector used to generate translational fusions. Similar constructs are performed on each of the two *pilC* genes from *N. meningitidis*. Antisera are generated against fusion proteins after their purification using conventional protocols. These antisera are extensively adsorbed with extracts of *E. coli* expressing alkaline phosphatase and β-galactosidase, and used in Western immunoblots and ELISA assays against a panel of *Neisseria gonorrhoeae* and *Neisseria meningitidis* strains. Antisera raised against fusion proteins carrying the major portion of PilC are also analyzed in Western blots using *E. coli* expressing fusion proteins containing only smaller regions of PilC.

The results of these studies should show which regions in PilC are immunodominant.

The *pilC1* and *pilC2* genes are highly homologous in their 5' ends whereas the homology is considerably less pronounced in the central and 3' region.

In addition, the entire *pilC2* gene from *N. gonorrhoeae* MS11(P^+) is cloned and sequenced. Algorithms are used to search for potential T-cell epitopes (amphipathic helical conformation) and β-cell epitopes. Polypeptides containing the predicted epitopes are tested to determine if they can prime mice for an enhanced immune response to PilC1 and PilC2.

Specific PilC antisera are used in immunoelectromicroscopy with piliated *Neisseria* cells as well as with purified pili to see if PilC is physically connected with the pilus fiber.

Neisseria is grown in the presence of different dilutions of $PilC^-$ specific antibodies. Bactericidal effect exerted by the antiserum, effects on piliation, and effects on bacterial attachment to corneal primary culture cells are monitored. Binding assays to epithelial cells are described in Tjia, K.F. et al. (1988), Graefe's Arch. Clin. Exp. Opthalmol. 226:341-345.

Example 8
Identification and characterization of genes located adjacent to *pilC*

The *pilC1* and *pilC2* loci are part of a larger duplication that extends both 5'- and 3'- of *pilC*. We know from our work with *E. coli* that strains may contain multiple gene clusters for the same class of pili. In one case we have shown that the only difference between two duplicated

gene clusters (*pap* and *prs*) resides in the adhesion genes such that each cluster gives rise to serologically identical pili binding to different cell surface receptors.

mTn*cm* mutagenesis in the region upstream and downstream of *pilC1* is performed to generate allelic replacements in the *pilC1* and *pilC2* regions on the chromosome. Since the two regions are highly homologous we expect to obtain for each insertion allelic replacements in either region. Double mutants are generated as before by isolating DNA from mutants carrying insertions in the *pilC2* region transforming P^+ variants carrying the same insertion in the *pilC1* region and select for transformants resistant to 30μg/ml of chloramphenicol. These double mutants are examined for piliation, pilins expression, and binding to corneal primary culture cells.

Example 9

Phase variation in gonococcal pili expression can be caused by frameshift mutations in *pilC*

If PilC is required for pilus formation, we would expect some P^- progeny arising from a P^+ clone to accumulate unassembled pilin in the absence of PilC. Nonpiliated (P^-) colonies were derived from $MS11_{mk}$(P^+), restreaked, and tested for the presence of PilC and pilin in immunoblots with the PilC and pili antisera. Five out of eight P^- clones did not produce detectable levels of PilC, but expressed the pilin subunit. The remaining three P^- clones expressed PilC but not pilin. The molecular mass of the pilin subunit was the same in the P^-, $PilC^-$ variants as in $MS11_{mk}$(P^+, $PilC^+$). However, the former in addition produced a protein reacting with the pili antiserum that was 16 kd in size. Since $MS11_{mk}$ only

contains one expression site for pilin we believe that this protein species represents a proteolytic degradation product of the pilin and may be identical to the S-pilin previously described (Haas et al., 1987). Several independent P^- clones were isolated from one P^-, $PilC^-$ clone. They all remained $PilC^-$ and retained expression of pilin. Piliated (P^+) revertants were also obtained from the same P^-, $PilC^-$ clone. These P^+ revertants occurred at about a tenfold lower frequency (10^{-4}) than P^- derivatives from a P^+ clone. All P^+ revertants from a P^-, $PilC^-$ clone had regained expression of PilC. All but one expressed a pilin with the same molecular weight as the nonpiliated parent. However, the low molecular weight pilin degradation product was much less abundant in the P^+, $PilC^-$ revertants. It was possible to obtain P^+ revertants from other P^-, $PilC^-$ clones as well, all of which expressed PilC.

The *pilE* gene from one set of PilC switches was PCR amplified and sequenced directly. The P^-, $PilC^-$, $pilin^+$ variant 8 carried eight amino acid changes in the pilin relative to the parental clone $MS11_{mk}$. The pilin sequence of the P^+, $PilC^-$ backswitcher 8:1 was identical to variant 8. Thus, the backswitching from P^- to P^+ colony morphology was not associated with any alteration in the pilus subunit protein implying that the change in colonial morphology was due to the switch in PilC expression.

Strain $MS11_{mk}(P^-)$, variants 8(P^-) and 8:1 (P^+) were also examined by transmission electron microscopy. Electron microscopy was performed with a JEOL 100CX microscope with 200-mesh copper grids coated with thin films of 2% Formvar. The bacterial colonies were carefully overlaid with buffer [10 μM Tris-HCl (pH 7.5), 10 μM magnesium chloride] and the cells were allowed to

sediment for 15 min on a grid. The grids were washed with water, negatively stained with 1% sodium silicotungstate (pH 7.0) and then washed again.

The electron micrographs showed that the MS11(P^+) parental cells were heavily piliated and pili were often seen to aggregate. In contrast most cells of variant 8(P^-) were nonpiliated. One or two pili were found on ~10% of these cells. All cells of variant 8:1 (P^+) were piliated, carrying ~10-40 fibers/cells. No aggregation of individual fibers was seen. These data confirm that the observed changes in colonial morphology reflect alterations in expression of pili. Therefore, phase variation of gonococcal pili may not only be caused by recombination events occurring in the *pilE* locus (Bergstrom et al, 1986; Swanson et al., 1986) but also by frameshift mutations in *pilC*.

Example 10

Immunogenicity of PilC

In order to predict a region of PilC which would have a high probability for antigenicity, residues 300 to 700 of the putative PilC1 protein encoded within *pilC1* were analyzed for antigen index, hydrophilicity, and hydrophobicity using standard computer-modelling methods. The analysis indicated that the PilC1 polypeptide fragment containing residues 300 to 700 would have several regions with a high antigen index, high hydrophilicity, and a high likelihood for location in an external domain.

The immunogenicity of a recombinant polypeptide expressed from the DNA encoding amino acids 300 to 700 was examined. The region of DNA encoding amino acid residues 300 to 700 was amplified by polymerase chain

reaction (PCR) using the following oligonucleotide primers.

5' GGC TAG GTG GCA TAT GAA AGA TAC CGG 3'

and

5' TTT GCA ATC GGG GAT CCT* C*A*G GTG TCT TTC 3'

These primers incorporate an NdeI and a BamHI restriction endonuclease site (indicated by the underlined nucleotides), respectively. A termination codon (indicated by the asterisks) was also incorporated. The PCR amplified DNA was then ligated into the vector pET3a (between the NdeI and BamHI sites). The recombinant vector pET3a is used in the inducible expression system described by Studier et al. (1990), using the protocol described therein. Strain BL21 (DE3) was transformed with the pET3a-*pilC* (300-700) vector, and the transformed strain used for the expression of the PilC (300-700) peptide.

The expression products after induction were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (PAGE) using an 8% acrylamide gel and the standard 25 mM Tris base / 250 mM glycine / pH 8.3 / 0.1% sodium dodecyl sulfate (SDS) electrophoresis buffer. After electrophoresis, the gel was fixed and then stained with Coomassie Blue according to standard protocols, and the production of PilC(300-700) was confirmed by the detection of the presence in the gels of an abundant, appropriately-sized peptide of approximately 46kD.

In order to detect the immunogenicity of the PilC(300-700) product, the region of the SDS-PAGE gel containing the PilC(300-700) polypeptide was excised from parallel unstained lanes, homogenized, and the protein eluted into a buffer of 0.1% Triton X-100 in water by

passive diffusion. Rabbits were given a priming intradermal injection of homogenized gel slices (containing approximately 500μg PilC(300-700) protein), followed 3 weeks later by a subcutaneous boost (of approximately 500 μg of eluted protein). An initial test serum was then collected after an additional 14 days. All of the test animals yielded a specific high titer antibody response. The antibodies induced by PilC(300-700) were immunologically reactive not only with that polypeptide (i.e., PilC1(300-700)), but also with native PilC1 and native PilC2.

The results demonstrate, inter alia, the following. PilC contains antigenic epitopes that can elicit a strong immunogenic response. At least some of the immunogenic epitopes are shared (cross-reactive) between PilC1 and PilC2, despite differences in primary amino acid sequence. The technique of subcloning discrete portions of the PilC protein under control of an inducible promoter allows mapping of antigenic epitopes. Sufficient quantities of specific oligopeptides of known antigenicity can be produced for use in screening the *in vivo* immune response after exposure to the intact pathogen.

The following listed materials are on deposit under the terms of the Budapest Treaty with the American Type Culture Collection (ATCC), 12301 Parklawn Dr., Rockville, Maryland 20852, and have been assigned the following Accession Numbers.

Description	ATCC No.	Deposit Date
pABJ03 in *E. coli* (DH5)	68519	Jan. 28, 1991
pABJ04 in *E. coli* (DH5)	68520	Jan. 28, 1991

Upon allowance and issuance of this application as a United States Patent, all restriction on availability of these deposits will be irrevocably removed; and access to the designated deposits will be available during pendency of the above-named application to one determined by the Commissioner to be entitled thereto under 37 CFR 1.14 and 35 USC 1.22. Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. The deposited materials mentioned herein are intended for convenience only, and are not required to practice the present invention in view of the descriptions herein, and in addition these materials are incorporated herein by reference.

Industrial Applicability

The invention, in the various manifestations disclosed herein, has many industrial uses, some of which are the following. The *pilC* DNAs may be used for the design of probes for the detection of *pilC* nucleic acids in samples. The probes derived from the DNAs may be used to detect *pilC* nucleic acids in, for example, chemical synthetic reactions. The polynucleotide probes are also useful in detecting viral nucleic acids in humans, and thus, may serve as a basis for diagnosis of pathogenic microorganisms containing type 4 pilin, for example, gonococcal and/or meningococcal infections in humans.

In addition to the above, the DNAs provided herein provide information and a means for synthesizing polypeptides containing epitopes of PilC. These polypeptides are useful in detecting antibodies to PilC antigens. A series of immunoassays the relevant

neisserial infection, based on recombinant polypeptides containing *pilC* epitopes are described herein, and will find commercial use in diagnosing diseases caused by these microorganisms. In addition, the polypeptides derived from the *pilC* DNAs disclosed herein will have utility as vaccines for treatment of infections caused by meningococci and gonococci.

The polypeptides derived from the *pilC* DNAs, besides the above stated uses, are also useful for raising anti-PilC antibodies. Thus, they may be used in vaccines against the relevant microorganisms. Moreover, the antibodies produced as a result of immunization with the polypeptides containing an immunoreactive PilC epitope are also useful as passive vaccines, or in the detection of the presence of PilC antigens in samples. Thus, they may be used to assay the production of polypeptides derived from PilC in chemical systems. The anti-PilC antibodies may also be used to monitor the efficacy of anti-neisserial agents in screening programs where these agents are tested in tissue culture systems. Another important use for anti-PilC antibodies is in affinity chromatography for the purification of PilC derived polypeptides. The purified PilC polypeptide preparations may be used in vaccines.

For convenience, the anti-PilC antibodies and polypeptides containing regions encoded in *pilC*, whether natural or recombinant, may be packaged into kits.

References cited herein

Badgasarian, M.M., et al., Gene (1983) 26:273-282.

Beard, M.K.M., et al., J. Bacteriol. (1990) 172:2601-2607.

Bergstrom, S., et al., Proc. Natl. Acad. Sci. USA (1986) 83:3890-3894.

Brinton, C.C., et al., in Brooks, G.F., et al. (eds.) Immunobiology of *Neisseria gonorrhoeae* (1978), American Society for Microbiology, Washington, D.C., pp. 155-178.

Edman, P., and Bregg, G., Eur. J. Biochem. (1967) 1:80-91.

Elleman, T.C., and Hoyne, P.A., J. Bacteriol. (1984) 160:1184-1187.

Elleman, T.C., et al., J. Bacteriol. (1986) 168:574-580.

Engstrom, A., et al., EMBO J. (1984) 1:2065-2070.

Fast, R., et al., Gene (1989) 85:227-231.

Fischer, S.H., and Rest, R.F., Infect. Immun. (1988) 56:1574-1579.

Frank-Kamenetskii, M.D., and Vologodskii, A.V., Nature (1984) 304:481-482.

Fürste, J.P., et al., Gene (1986) 48:119-131.

Gill, D.R., et al., Mol. Gen. Genet. (1986) 205:134-145.

Gotschlich, E.C., et al., J. Exp. Med. (1986) 164:868-881.

Haas, R., and Meyer, T., Cell (1986) 44:107-115.

Haas, R., et al., Proc. Natl. Acad. Sci. USA (1987) 84:9079-9083.

Hagblom, P., et al., Nature (1985) 315:156-158.

Hanahan, D., in Glover, D.M. (ed.), DNA Cloning: A Practical Approach (1985), Vol. I., pp. 109-129 (IRL Press, Oxford).

Hanson, M.S., and Brinton, C.C., Nature (1988) 322:265-268.

Hill, S.A., et al., Mol. Microbiol. (1990) 4:1341-1352.

Hultgren, S.J., et al., Proc. Natl. Acad. Sci. USA (1989) 86:4357-4361.

Kellogg, D.S., et al., J. Bacteriol. (1968) 96:596-605.

Laemmli, U.K., Nature (1970), 227:680-685.

Lindberg, F.P., et al., Nature (1987) 328:84-87.

Lindberg et al., J. Bacteriol. (1989) 171:6052-6058.

Marrs, C.F., et al., J. Bacteriol. (1985) 163:132-139.

Mattick, J.S., et al., J. Bacteriol. (1987) 169:33-41.

McKern, N.M., et al., FEBS Lett. (1983) 164:149-153.

Meyer, T., et al., Cell (1982) 30:45-52.

Meyer et al., Proc. Natl. Acad. Sci. USA (1984) 81:6110-6114.

Moch, T., et al., Proc. Natl. Acad. Sci. USA (1987) 84:3462-3466.

Mool, F.R., et al., J. Bacteriol. (1983) 154:41-49.

Murphy, G.L., et al., Cell (1989) 56:589-547.

Norgren, M., et al., Mol. Microbiol. (1987) 1:169-178.

Norlander, L., et al., J. Bacteriol. (1981) 145:788-795.

Norquist, A., et al., FEMS Microbiol. Lett. (1978) 4:71-75.

Orndorff, P.E., and Falkow, S., J. Bacteriol. (1984) 159:736-744.

Panyutin, I.G., et al., Proc. Natl. Acad. Sci. USA (1990) 87:867-870.

Pasloske, B.L., et al., FEBS Lett. (1985) 183:408-412.

Pearce, W.C., and Buchanan, T.M., J. Clin. Invest. (1978) 61:931-943.

Sanger, F., et al., Proc. Natl. Acad. Sci. USA (1977) 74:5463-5467.

Sanger, F., et al., J. Mol. Biol. (1980) 143:161-178.

Segal. E., et al., Cell (1985) 40:293-300.

Seifert, H.S., et al., Proc. Natl. Acad. Sci. USA (1986) 83:735-739.

Seifert, H.S., et al., J. Bacteriol. (1990) 172:40-46.

Shaw, C.E., and Taylor, R.K., Infect. Immun. (1990) 58:3042-3049.

Southern, E.M., J. Mol. Biol. (1975) 98:503-517.

Stern, A., et al., Cell (1986) 47:61-71.

Stibitz, S., et al., Nature (1989) 338:266-269.

Stoker, N.G., et al., in Hames, B.D., and Higgin, S.J. (eds.), Transcription and Translation: A Practical Approach (1984) (IRL Press, Oxford).

Studier et al., Methods in Enzymology (1990) 185:60-89.

Swanson, J., J. Exp. Med. (1973) 137:571-589.

Swanson, J., et al., J. Exp. Med. (1985) 162:729-744.

Swanson, J., et al., Cell (1986) 47:267-276.

Taha, M.K., et al., EMBO J. (1988) 7:4367-4378.

Taylor, T., et al., Vaccine (1988) 6:151-154.

Thompson, R., and Achtman, M., Mol. Gen. Genet. (1978) 165:295-304.

Towbin, H., et al., Proc. Natl. Acad. Sci. USA (1979) 76:4350-4354.

von Heinje, G., Eur. J. Biochem. (1983) 133:17-21.

Watt, P.J., and Ward, M.E., in Beachey, E.H. (ed.), Bacterial Adherence (1980) Series B, Vol. 6, pp. 253-288, Chapman & Hall, London.

Weiser, J.N., et al., Cell (1989) 59:657-665.

Willems, R., et al., EMBO J. (1990) 9:2803-2809.

Yanis-Perron, C., et al., Gene (1985) 33:103-119.

Broach (1981) in Molecular Biology of the Yeast Saccharomyces, Vol. 1, p.445, Cold Spring Harbor Press.

Broach et al., Meth. Enz. (1983) 101:307.

Catty (1988), ANTIBODIES, Volume 1: A PRACTICAL APPROACH (IRL Press).

Chaney et al. (1986), Cell and Molecular Genetics 12:237.

Chakrabarti et al. (1985), Mol. Cell Biol. 5:3403.

Cohen (1972), Proc. Natl. Acad. Sci. USA 69:2110.

Glennie et al. (1982), Nature 295:712.

Grunstein and Hogness (1975), Proc. Natl. Acad. Sci. USA 73:3961.

Grych et al. (1985), Nature 316:74.

Hammerling et al. (1981), MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS.

Hess et al. (1968), J. Adv. Enzyme Reg 7:149.

Hinnen et al. (1978), Proc. Natl. Acad. Sci. 75:1929.

Hitzeman et al. (1980), J. Biol. Chem. 255:2073.

Holland et al. (1978), Biochemistry 17:4900.

Holland (1981), J. Biol. Chem. 256:1385.

Holland and Holland (1980), J. Biol. Chem. 255:2596.

Hunyh, T.V. et al. (1985) in DNA CLONING TECHNIQUES; A PRACTICAL APPROACH (D. Glover, Ed., IRL Press, Oxford, U.K.) pp. 49-78.

Immun. Rev. (1982) 62:185.

Ito et al. (1984), Agric. Biol. Chem. 48:341.

Kennett et al. (1980) MONOCLONAL ANTIBODIES.

Laemmli (1970), Nature 227:680.

Luckow and Summers (1989), Virology 17:31.

Mackett et al. (1984), J. Virol. 49:857.

Maniatis, T., et al. (1982) MOLECULAR CLONING; A LABORATORY MANUAL (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Maniatis et al. (1989), MOLECULAR CLONING; A LABORATORY MANUAL, Second Edition (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.).

Mayer and Walker, eds. (1987), IMMUNOCHEMICAL METHODS IN CELL AND MOLECULAR BIOLOGY (Academic Press, London).

Maxam et al. (1980), Methods in Enzymology 65:499.

MacNamara et al. (1984), Science 226:1325.

Messing et al. (1981), Nucleic Acids Res. 9:309.

Messing (1983), Methods in Enzymology 101:20-37.

METHODS IN ENZYMOLOGY (Academic Press).

Moss (1987) in GENE TRANSFER VECTORS FOR MAMMALIAN CELLS (Miller and Calos, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), p. 10.

Neurath et al. (1984), Science 224:392.

Saiki et al. (1986), Nature 324: 163.

Saiki et al. (1988), Science 239:487.

Sanger et al. (1977), Proc. Natl. Acad. Sci. USA 74:5463.

Setlow, ed. (1988), GENETIC ENGINEERING. Vol. 10, p. 195-219 (Plenum Publishing Co., N.Y.

Schreier, M., et al. (1980) HYBRIDOMA TECHNIQUES

Scopes (1984), PROTEIN PURIFICATION, PRINCIPLES AND PRACTICE, Second Edition (Springer-Verlag, N.Y.).

Tsu and Herzenberg (1980), in SELECTED METHODS IN CELLULAR IMMUNOLOGY (W.H. Freeman and Co.) pp. 373-391.

Vytdehaag et al. (1985), J. Immunol. 134:1225.

Valenzuela, P., et al. (1982), Nature 298:344.

Valenzuela, P., et al. (1984), in HEPATITIS B (Millman, I., et al., ed, Plenum Press) pp. 225-236.

Warner (1984), DNA 3:401.

Wu and Grossman (1987), METHODS IN ENZYMOLOGY Vol. 154, RECOMBINANT DNA, Part E.

Wu (1987), METHODS IN ENZYMOLOGY Vol 155, RECOMBINANT DNA, part F.

Zoller (1982), Nucleic Acids Res. 10:6487.

CLAIMS

1. A recombinant polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

2. The recombinant polynucleotide of claim 1, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

3. The recombinant polynucleotide of claim 2, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

4. The recombinant polynucleotide of claim 2, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

5. A vector comprised of a recombinant polynucleotide, wherein the recombinant polynucleotide is selected from the group consisting of the recombinant polynucleotide of claim 1, the recombinant polynucleotide of claim 2, the recombinant polynucleotide of claim 3, and the recombinant polynucleotide of claim 4.

6. A host cell transformed with the vector of claim 5.

7. A recombinant expression system comprising a polynucleotide encoding a polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*, wherein the polynucleotide is operably linked to a control sequence compatible with a desired host.

8. The recombinant expression system of claim 7, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

9. The recombinant expression system of claim 8, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

10. The recombinant expression system of claim 8, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

11. A cell transformed with a recombinant expression system, wherein the expression system is selected from the recombinant expression system of claim 7, the recombinant expression system of claim 8, the recombinant expression system of claim 9, and the recombinant expression system of claim 10.

12. A polypeptide produced by the cell of claim 11.

13. A purified polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

14. A purified polypeptide according to claim 13, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

15. A purified polypeptide according to claim 14, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

16. A purified polypeptide according to claim 14, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

17. A recombinant polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

18. A recombinant polypeptide according to claim 17, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

19. A recombinant polypeptide according to claim 18, wherein the immunoreactive epitope is encoded in *pilC2* of *Neisseria gonorrheae*.

20. A recombinant polypeptide according to claim 18, wherein the immunoreactive epitope is encoded in *pilC1* of *Neisseria gonorrheae*.

21. A method of preparing a recombinant polypeptide comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*, the method comprising:

a. providing a host cell according to claim 11;

b. incubating the cell under conditions which allow expression of the recombinant polypeptide; and

c. isolating the polypeptide.

22. A vaccine composition for the treatment of *Neisseria infection*, comprised of a pharmaceutically acceptable excipient and of an effective amount of a recombinant polypeptide, wherein the polypeptide is

comprised of an immunoreactive epitope of a protein encoded in *pilC* of *Neisseria*.

23. The vaccine composition of claim 22, wherein the immunoreactive epitope is encoded in *pilC* of *Neisseria gonorrheae*.

24. A polypeptide affixed to a solid substrate, wherein the polypeptide is selected from the group consisting of the polypeptide of claim 12, the polypeptide of claim 13, the polypeptide of claim 14, the polypeptide of claim 15, the polypeptide of claim 16, the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, and the polypeptide of claim 20.

25. An immunoassay for detection of anti-*Neisseria* antibodies comprising:

(a) providing a sample suspected of containing anti-*Neisseria* antibodies;

(b) providing an antigen, wherein the antigen is a polypeptide selected from the group consisting of the polypeptide of claim 12, the polypeptide of claim 13, the polypeptide of claim 14, the polypeptide of claim 15, the polypeptide of claim 16, the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, the polypeptide of claim 20; and

(c) incubating the sample of (a) with the antigen of (b) under conditions which allow the formation of antibody-antigen complexes; and

(d) detecting the presence of anti-*Neisseria* antibody-antigen complexes formed in (c), if any.

26. A composition comprised of a polypeptide, wherein the polypeptide is selected from the group

consisting of the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, and the polypeptide of claim 20.

27. A composition comprised of purified polyclonal anti-PilC antibodies, wherein the PilC is of *Neisseria*.

28. A composition comprised of a monoclonal antibody directed against an immunoreactive epitope encoded in *pilC* of *Neisseria*.

29. An immunoassay for detection of an antigen encoded in *pilC* of *Neisseria* comprising:

(a) providing a sample suspected of containing an antigen encoded in *pilC* of *Neisseria*;

(b) providing a composition comprised of antibodies directed against the antigen encoded in *pilC* of *Neisseria*, wherein the composition is selected from the group of compositions of claim 27 and claim 28;

(c) reacting the sample of (a) with the antibody containing composition of (b) under conditions which allow the formation of anti-PilC antibody-antigen complexes;

(d) detecting anti-PilC antibody-antigen complexes formed in (c), if any.

30. A kit for analyzing samples for the presence of anti-PilC antibodies comprising:

(a) an antigen packaged in a suitable container, wherein the antigen is a polypeptide selected from the group consisting of the polypeptide of claim 12, the polypeptide of claim 13, the polypeptide of claim 14, the polypeptide of claim 15, the polypeptide of claim 16,

the polypeptide of claim 17, the polypeptide of claim 18, the polypeptide of claim 19, and the polypeptide of claim 20;

(b) a buffer used in the performance of the analysis, packaged in a suitable container; and

(c) instructions on the performance of the analysis which uses the antigen of (a) and the buffer of (b).

31. A kit for analyzing samples for the presence of an antigen comprised of an immunoreactive epitope encoded in *pilC* of *Neisseria* comprising:

(a) a composition comprised of antibodies directed against the antigen comprised of an immunoreactive epitope encoded in *pilC* of *Neisseria*, wherein the composition is selected from the group of compositions of claim 27 and claim 28, wherein the composition is packaged in a suitable container;

(b) a buffer used in the performance of the analysis, packaged in a suitable container; and

(c) instructions for performing the analysis.

32. A method for producing antibodies to PilC of *Neisseria* comprising administering to an individual a composition comprised of an isolated immunogenic polypeptide containing a PilC epitope in an amount sufficient to produce an immune response to the PilC epitope.

33. An oligomer capable of hybridizing to a sequence in *pilC* of *Neisseria*, wherein the oligomer is comprised of a *pilC* sequence complementary to at least about 6 contiguous nucleotides of *pilC*.

34. An oligomer according to claim 33, wherein *pilC* is of *Neisseria gonorrheae*.

35. A process for detecting a *pilC* sequence in an analyte strand, wherein the *pilC* sequence comprises a selected target region, the process comprising:

(a) providing a sample comprised of an analyte strand suspected of containing a selected target *pilC* sequence;

(b) providing an oligomer capable of hybridizing to the target *pilC* sequence, wherein the oligomer is comprised of a *pilC* targeting sequence complementary to at least about 6 contiguous nucleotides of *pilC*;

(c) incubating the sample of (a) with the oligomer of (b) under conditions which allow specific hybrid duplexes to form between the targeting sequence and the target sequence; and

(d) detecting hybrids formed between the target sequence, if any, and the oligomer.

36. The process of claim 35 which further comprises:

(a) providing a set of oligomers which are primers for a polymerase chain reaction (PCR) method and which flank the target region; and

(b) amplifying the target region via the PCR method.

37. A kit for detecting a *pilC* sequence in an analyte strand comprising:

(a) the oligomer of claim 33, packaged in a suitable container;

(b) a suitable buffer, packaged in a suitable container; and

(c) instructions for performing the detection.

38. A recombinant polynucleotide comprising a DNA sequence of at least 8 contiguous nucleotides from *pilC*, wherein the *pilC* sequence is selected from the group of sequences shown in Figure 3, Figure 6, and Figure 7.

39. A method of treating an individual for a *Neisseria* infection comprising administering to the individual antibodies produced according to claim 32, wherein the antibodies are administered in an amount effective to prevent the pathology of the infection.

40. An immunoassay for detection of anti-*Neisseria* antibodies comprising:

(a) providing a sample suspected of containing anti-*Neisseria* antibodies;

(b) providing an antigen, wherein the antigen is the polypeptide of claim 24;

(c) incubating the sample of (a) with the antigen of (b) under conditions which allow the formation of antibody-antigen complexes; and

(d) detecting the presence of anti-*Neisseria* antibody-antigen complexes formed in (c), if any.

41. A kit for analyzing samples for the presence of anti-PilC antibodies comprising:

(a) an antigen packaged in a suitable container, wherein the antigen is a polypeptide from claim 24;

(b) a buffer used in the performance of the anaylsis, packaged in a suitable container; and

(c) instructions on the performance of the analysis which uses the antigen of (a) and the buffer of (b).

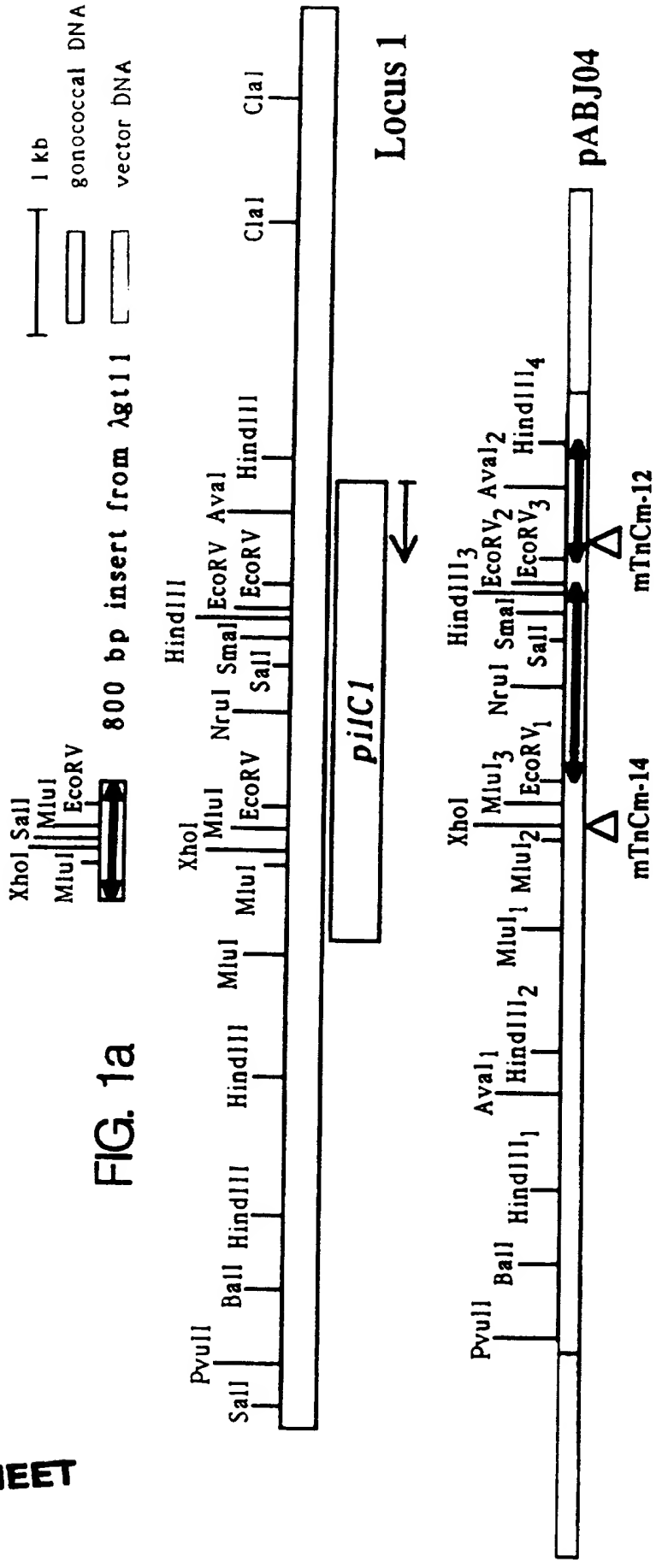

FIG. 1a

SUBSTITUTE SHEET

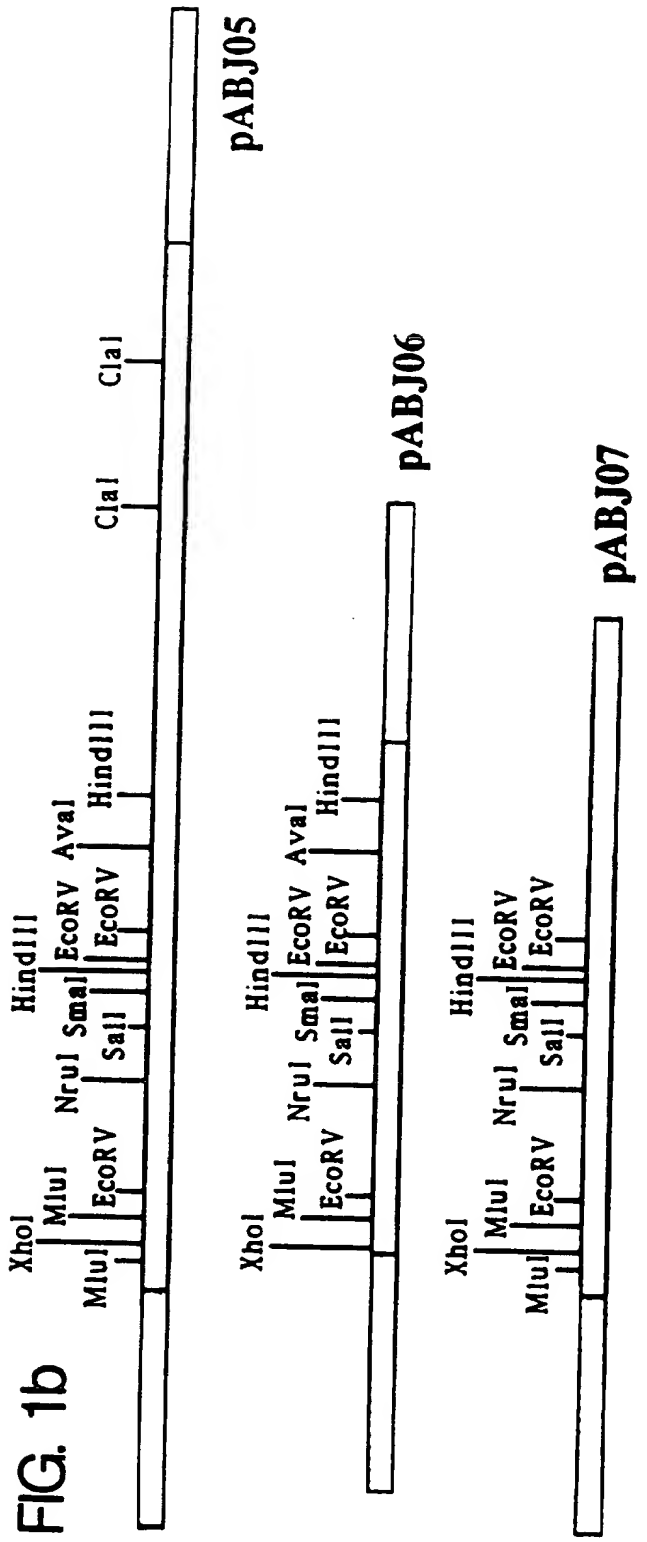

FIG. 1b

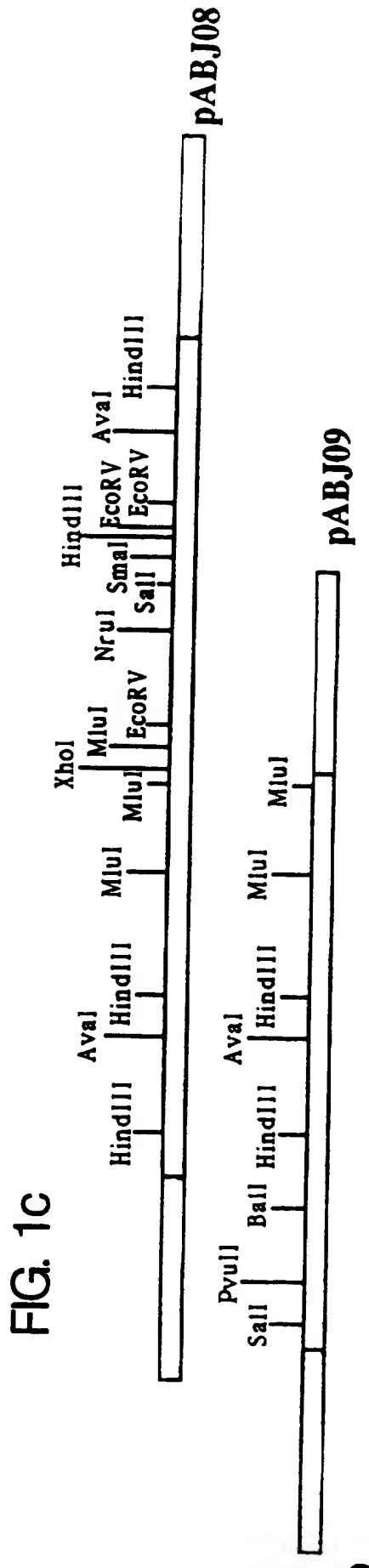

FIG. 1C

```
         90                110               130               150               170
CGTCCCGCGAAGGCAAACTTAAGGAATAAAATATGAATAAAACTTTGAAACGGCAGGTTTTCCGCCATACCGCGCTTTATGCCGCCATCTTGATGTTTTC
                          frame 1 MetAsnLysThrLeuLysArgGlnValPheArgHisThrAlaLeuTyrAlaAlaIleLeuMetPheSe

         190               210               230               250               270
CCATACCGGCGGGGGGGGGGGCGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGAAACCAGCCCGAGGTAAAGCAGAATGTGCCATCTT
rHisThrGlyGlyGlyGlyGlyArgTrpArgLysProIleAsnThrLeuLeuSerEnd
              frame 2 MetAlaGlnThrHisGlnTyrAlaIleIleMetAsnGluArgAsnGlnProGluValLysGlnAsnValProSerS
```

1	2	3	4	5	6	7	8	9	10
?	?	?	Lys	Tyr	Ala	Ile	Ile	Met	Asn

FIG. 2

```
10                  30                  50
GATCCGCCCCGGTGCTTGGGCGCCTTAGGGAACCGTTCCCTTTGAGCCGGGGCGGGGCAAC
70                  90                  110
GCGTACCGGTTTTTGTTAATCCGCTATAAAGGCGGGCTATAGGGTAGGCTTCATCCTG
130                 150                 170
CCAATCTCACTGAATCCGTCAATTTCCGCAATTCAATTAAATACCGTCAAACCGATGCCG
190                 210                 230
TCATTCCGCCGCAGGCGGGAATCCGGACCGGTCGGGCATCTGCGGCGGTTTGCTAAAAAAC
250                 270                 290
GCTTTACCGTGATAAGTGCGCAAAGTTAAAATGGGGAGGTAAGCTTTTCAATCAGCAATC
310                 330                 350
CGGCGGGCGCGGAATCGGGCGGTTTACCGAACCCCGGCGTTCGCGGCGCCCGTCCCGCGA
370                 390                 410
AGGCAAACTTAAGGAATAAAATATGAATAAAACTTGAAACGGCAGGTTTTCCGCCATAC
430                 450                 470
CGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGCGATGGCG
490                 510                 530
CAAACCCATCAATACGCTATTATCATGAACGAGCGAAACCAGCCCGAGGTAAAGCAGAAT
550                 570                 590
GTGCCATCTTCAATAAAGGACAAAGACAGGAGGCGCGAATATACTTATTATACGCACAGA
610                 630                 650
ACAGGAGGAGGCTCTGTCTCATTCAACAATAACGATACCCTTGTTTCCCAACAAAGCGGT
```

FIG. 3-1a

```
670                 690                 710
ACTGCCGTTTTTGGCACAGCCACCTACCTGCCGCCCTACGGCAAGGTTTCCGGTTTTGAT
730                 750                 770
GCCGTCGCTCTGAAAGAGCGCAACAATGCCGTTGATTGGATTCGTACCACCCGCATCGCG
790                 810                 830
CTGGCAGGCTACTCCTACATCGACGTCATATGCAGAAGCTACACACAGGCTGTCCCAAACTT
850                 870                 890
GTCTATAAAACCCGATTTACCTTCGGTCAACAAGGGTTGAAAAGAAAGGCAGGCAGCAAG
910                 930                 950
CTGGATATATACGAAGACAAAAGCCGCGAAAATTCGCCCATTTACAAATTGTCGGATTAT
970                 990                 1010
CCTTGGTTGGGCGTATCTTTCAATTTGGGCAGCGAGAATACCGTCCAAAATAGCAAATTA
1030                1050                1070
TTCAACAAATTGATATCTTCTTTTAGAGAAGGCAATAATAATCAAACCATCGTCTCTACG
1090                1110                1130
ACAGAAGGCAACCCTATTTCCCTTGGCGACCGGCAGCGCGAACATACCGCCGTGGCCTAT
1150                1170                1190
TATCTGAACGCCAAACTGCACCTGCTGGACAAAAAAGGGATTGAAGATATCGCCCAAGGC
1210                1230                1250
AAAATAGTGGATTTGGGTATCTTGAAACCGCACGTCGAGACGACAGGACGAAGCTTGCTA
```

FIG. 3-1b

```
1270        1290        1310
GATTTTTGGGCTAGGTGGGACATTAAAGATACCGGGCAGATTCCGGTCAAGCTCGGCCTG
   1330        1350        1370
CCGCAAGTCAAAGCAGGCCGCTGCACCAACAAACCGAACCCCAATAATACCAAAGCC
   1390        1410        1430
CCTTCGCCGGCACTGACCGCCCCCCGCGCTGTGGTTCGGACCCGGGCAAGATGGTAAGGCG
   1450        1470        1490
GAGATGTATTCCGCTTCGGTTTCCACCTACCCCGACAGTTCGAGCAGCCGCATCTTCCTC
   1510        1530        1550
CAAGAGCTGAAAACTCAAACCGAACCCGGCAAACCCGGCCGCTATTCCCTCAAATCTTTG
   1570        1590        1610
AATGATGGTGAGATTAAAAGTCGACAGCCGAGTTTCAACGGGCGGCAAACAATCATCCGA
   1630        1650        1670
TTGGATGACGGCGTACATTTGATCAAACTGAATGGAAGCAAGGATGAGGTCGCCGCTTTT
   1690        1710        1730
GTCAATTTAAATGGAAACAACACCGGCAAAAACGACACTTTCGGCATTGTTAAGGAAGCG
   1750        1770        1790
AACGTCAATCTTGACGCCGACGAGTGGAAAAAGTGCTGCTGCCTTGGACGGTTCGGGGT
   1810        1830        1850
CCCGATAATGACAATAAATTTAAATCAATTAACCAAAAACCAGAAAAATACAGCCAAAGA
```

FIG. 3-1C

1870 1890 1910
TACCGCATCCGCGACAACAACGGCAATCGCGATTTGGGCGACATCGTCAACAGCCCGATT
1930 1950 1970
GTCGCGGTCGGCGGGTATTTGGCAACCGCCGCGAACGACGGGATGGTGCATATCTTCAAA
1990 2010 2030
AAAAACGGCGGCAGTGATGAACGCAGCTACAATCTGAAGCTCAGCTACATCCCCGGCACG
2050 2070 2090
ATGCCGCGCAAGGATATTCAAAGCCAAGAATCCACCCTTGCCAAAGAGCTGCGCGCCTTT
2110 2130 2150
GCCGAAAAAGGCTATGTGGGCGACCGCTACGGCGTGGACGGCGGCTTTGTCTTGCGCCAA
2170 2190 2210
GTCGAACTGAGCGGGCAAAAACACGTGTTTATGTTCGGCGCGATGGGTTTTGGCGGCAGG
2230 2250 2270
GGCGCGTATGCCTTGGATTTAAGCAAAATCAACGGAAATTATCCGGCCGCGCCCCCCTG
2290 2310 2330
TTTGATGTCAAAGATGGCGATAATAACGGCAAAAATCGCGTGAAAGTGGAATTAGGCTAC
2350 2370 2390
ACCGTCGGTACGCCGCAAATCGGCAAAATCCGCAACGGCAAATACGCCGCCTTCCTCGCC
2410 2430 2450
TCCGGTTATGCGGCTAAAAAAATTGACGACTCAACAAATAAAACCGCGCTGTATGTATAT

FIG. 3-2a

```
2470        2490        2510
GATTTGAAAGACACCTTAGGTACGCCGATTGCAAAAATCGAAGTGAAGGACGGCAAAGGC
2530        2550        2570
GGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACGGTCGATATCGCCTAT
2590        2610        2630
GCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAATTCCGATTCTAGTAAA
2650        2670        2690
TGGTCTGCAAAGGTTATTTCGAAGGCGACAAGCCGATTACCTCCGCGCCCGCCGTTTCC
2710        2730        2750
CGACTGGCAGACAAACGCGTCGTCATCTTCGGTACGGGCAGCGATTTGACCGAAGATGAT
2770        2790        2810
GTACTGAATACGGGCGAACAATATATTTACGGTATCTTTGACGACCGATAAGGGACGGTT
2830        2850        2870
AAGGTAACGGTACAAAACGGCACGGCAGGCGGGCTGCTCGAGCAACACCTTACTCAGGAA
2890        2910        2930
AATAAAACATTATTCCTGAACAAGAGATCCGACGGTTCGGGCAGCAAGGGCTGGGCGGTG
2950        2970        2990
AAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACCGTGGTATTGCGTACCGCCTTC
3010        3030        3050
GTAACCATCCGCAAATATAACGACGGCGGCTGCGGCGCGGAAACCGCCATTTTGGGCATC
```

FIG. 3-2b

```
      3070                3090                3110
AATACCGCCGACGGCGGCGCATTGACTCCGAGAAGCGCGCCCGATTGTGCCGGATCAC
      3130                3150                3170
AATTCGGTTGCGCAATATTCCGGCCATAAGACAACCTCCAAAGGCAAATCCATCCCTATA
      3190                3210                3230
GGTTGTATGGACAAAGACGGTAAAACCGTCTGCCCGAACGGATATGTTTACGACAAGCCG
      3250                3270                3290
GTTAATGTGCGTTATCTGGATGAAACGGAAACAGACGGATTTTCAACGACGGCGGACGGC
      3310                3330                3350
GATGCGGGCGGCAGCGGTATAGACCCCGCCGGCAGGCGTCCCGGCAAAAACAACCGCTGC
      3370                3390                3410
TTCTCCAAAAAAGGGGTGCGCACCCTGCTGATGAACGATTTGGACAGCTTGGATATTACC
      3430                3450                3470
GGCCCGATGTGCGGTATCAAACGCTTAAGCTGGCGCGAAGTCTTCTTCTGACCGGCCTGC
      3490                3510                3530
GCGGCCGGTTTTTCCGCAAATGCCGTCCGAAAGGCCTTCGGACGGCATTTTTTTTGCGTTT
      3550
TTCGGGAGGGGGGCGGCAAATGAAACG
```

FIG. 3-2C

FIG. 4-1a

```
          10                                30                                50
GATCCGCCCGGTGCTTGGGCGCCTTAGGGAACCGTTCCCTTTGAGCCGGGGCGGGGCAAC
          70                                90                               110
GCGTACCGGTTTTTGTTAATCCGCTATAAAAGGCGGGCTATAGGGTAGGCTTCATCCTGC
         130                               150                               170
CAATCTCACTGAATCCGTCAATTTCCGCAATTCAATTAAATACCGTCAAACCGATGCCGT
         190                               210                               230
CATTCCGCGCAGGCGGGAATCCGGACCGGTCGGGCATCTGCGGCGGTTTGCTAAAAAACG
         250                               270                               290
CTTTACCGTGATAAGTGCGCAAAGTTAAAATGGGGAGGTAAGCTTTTCAATCAGCAATCC
         310                               330                               350
GGCGGGCGCGGAATCGGGCGGTTTACCGAACCCCGGCGTTCGCGGCGCCCGTCCCGCGAA
         370                               390                               410
GGCAAACTTAAGGAATAAAATATGAATAAAACTTTGAAACGGCAGGTTTTCCGCCATACC
                       M  N  K  T  L  K  R  Q  V  F  R  H  T
         430                               450                               470
GCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGCGATGGCG
A  L  Y  A  A  I  L  M  F  S  H  T  G  G  G  G  G  A  M  A
         490                               510                               530
CAAACCCATCAATACGCTATTATCATGAACGAGCGAAACCAGCCCGAGGTAAAGCAGAAT
Q  T  H  Q  Y  A  I  I  M  N  E  R  N  Q  P  E  V  K  Q  N
```

FIG. 4-1b

```
          550                   570                   590
GTGCCATCTTCAATAAAGGACAAAGACAGGAGGCGCGAATATACTTATTATACGCACAGA
V  P  S  S  I  K  D  K  D  R  R  R  E  Y  T  Y  Y  T  H  R
          610                   630                   650
ACAGGAGGAGGCTCTGTCTCATTCAACAATAACGATACCCTTGTTTCCCAACAAAGCGGT
T  G  G  G  S  V  S  F  N  N  N  D  T  L  V  S  Q  Q  S  G
          670                   690                   710
ACTGCCGTTTTTGGCACAGCCACCTACCTGCCGCCCTACGGCAAGGTTTCCGGTTTTGAT
T  A  V  F  G  T  A  T  Y  L  P  P  Y  G  K  V  S  G  F  D
          730                   750                   770
GCCGTCGCTCTGAAAGAGCGCAACAATGCCGTTGATTGGATTCGTACCACCCGCATCGCG
A  V  A  L  K  E  R  N  N  A  V  D  W  I  R  T  T  R  I  A
          790                   810                   830
CTGGCAGGCTACTCCTACATCGACGTCATATGCAGAAGCTACACAGGCTGTCCCAAACTT
L  A  G  Y  S  Y  I  D  V  I  C  R  S  Y  T  G  C  P  K  L
          850                   870                   890
GTCTATAAAACCCGATTTACCTTCGGTCAACAAGGGTTGAAAAGAAAGGCAGGCAGCAAG
V  Y  K  T  R  F  T  F  G  Q  Q  G  L  K  R  K  A  G  S  K
          910                   930                   950
CTGGATATATACGAAGACAAAAGCCGCGAAAATTCGCCCATTTACAAATTGTCGGATTAT
L  D  I  Y  E  D  K  S  R  E  N  S  P  I  Y  K  L  S  D  Y
```

FIG. 4-1c

```
970                                     990                                     1010
CCTTGGTTGGGCGTATCTTTCAATTTGGGCAGCGAGAATACCGTCCAAAATAGCAAATTA
P   W   L   G   V   S   F   N   L   G   S   E   N   T   V   Q   N   S   K   L
1030                                    1050                                    1070
TTCAACAAATTGATATCTTCTTTTAGAGAAGGCAATAATAATCAAACCATCGTCTCTACG
F   N   K   L   I   S   S   F   R   E   G   N   N   N   Q   T   I   V   S   T
1090                                    1110                                    1130
ACAGAAGGCAACCCTATTTCCCTTGGCGACCGGCAGCGCGAACATACCGCCGTGGCCTAT
T   E   G   N   P   I   S   L   G   D   R   Q   R   E   H   T   A   V   A   Y
1150                                    1170                                    1190
TATCTGAACGCCAAACTGCACCTGCTGGACAAAAAAGGGATTGAAGATATCGCCCAAGGC
Y   L   N   A   K   L   H   L   L   D   K   K   G   I   E   D   I   A   Q   G
1210                                    1230                                    1250
AAAATAGTGGATTTGGGTATCTTGAAACCGCACGTCGAGACGACAGGACGAAGCTTGCTA
K   I   V   D   L   G   I   L   K   P   H   V   E   T   T   G   R   S   L   L
```

```
                1270                1290                1310
GATTTTTGGGCTAGGTGGGACATTAAAGATACCGGGCAGATTCCGGTCAAGCTCGGCCTG
D  F  W  A  R  W  D  I  K  D  T  G  Q  I  P  V  K  L  G  L
          1330                1350                1370
CCGCAAGTCAAAGCAGGCCGCTGCACCAACAAACCGAACCCCAATAATACCAAAGCC
P  Q  V  K  A  G  R  C  T  N  K  P  N  P  N  N  N  T  K  A
          1390                1410                1430
CCTTCGCCCGGCACTGACCGCCCCCGCGCTGTGGTTCGGACCCGGGCAAGATGGTAAGGCG
P  S  P  A  L  T  A  P  A  L  W  F  G  P  G  Q  D  G  K  A
          1450                1470                1490
GAGATGTATTCCGCTTCGGTTCCACCTACCCCGACAGTTCGAGCAGCCGCATCTTCCTC
E  M  Y  S  A  S  V  S  T  Y  P  D  S  S  S  S  R  I  F  L
          1510                1530                1550
CAAGAGCTGAAAACTCAAACCGAACCCGGCAAACCCGGCCGCTATTCCCTCAAATCTTTG
Q  E  L  K  T  Q  T  E  P  G  K  P  G  R  Y  S  L  K  S  L
          1570                1590                1610
AATGATGGTGAGATTAAAAGTCGACAGCCGAGTTTCAACGGGCGGCAAACAATCATCCGA
N  D  G  E  I  K  S  R  Q  P  S  F  N  G  R  Q  T  I  I  R
          1630                1650                1670
TTGGATGACGGCGTACATTTGATCAAACTGAATGGAAGCAAGGATGAGGTCGCCGCTTTT
L  D  D  G  V  H  L  I  K  L  N  G  S  K  D  E  V  A  A  F
          1690                1710                1730
GTCAATTTAAATGGAAACAACACCGGCAAAAACGACACTTTCGGCATTGTTAAGGAAGCG
V  N  L  N  G  N  N  T  G  K  N  D  T  F  G  I  V  K  E  A
```

FIG. 4-2a

```
1750                1770                1790
AACGTCAATCTTGACGCCGACGAGTGGAAAAAGTGCTGCCTTGCCTTGGACGGTTCGGGGT
N  V  N  L  D  A  D  E  W  K  K  V  L  L  P  W  T  V  R  G
      1810                1830                1850
CCCGATAATGACAATAAATTTAAATCAATTAACCAAAAACCAGAAAAATACAGCCAAAGA
P  D  N  D  N  K  F  K  S  I  N  Q  K  P  E  K  Y  S  Q  R
      1870                1890                1910
TACCGCATCCGCGACAACAACGGCAATCGCGATTTGGGCGACATCGTCAACAGCCCGATT
Y  R  I  R  D  N  N  G  N  R  D  L  G  D  I  V  N  S  P  I
      1930                1950                1970
GTCGCGGTCGGCGGGTATTTGGCAACCGCCGCGAACGACGGGATGGTGCATATCTTCAAA
V  A  V  G  G  Y  L  A  T  A  A  N  D  G  M  V  H  I  F  K
      1990                2010                2030
AAAACGGCGGCAGTGATGAACGCAGCTACAATCTGAAGCTCAGCTACATCCCCGGCACG
K  N  G  G  S  D  E  R  S  Y  N  L  K  L  S  Y  I  P  G  T
      2050                2070                2090
ATGCCGCGCAAGGATATTCAAAGCCAAGAATCCACCCTTGCCAAAGAGCTGCGCGCCTTT
M  P  R  K  D  I  Q  S  Q  E  S  T  L  A  K  E  L  R  A  F
      2110                2130                2150
GCCGAAAAAGGCTATGTGGGCGACCGCTACGGCGTGGACGGCGGCTTTGTCTTGCGCCAA
A  E  K  G  Y  V  G  D  R  Y  G  V  D  G  G  F  V  L  R  Q
```

FIG. 4-2b

FIG. 4-2c

```
      2170                2190                2210
GTCGAACTGAGCGGGCAAAAACACGTGTTTATGTTCGGCGCGATGGGTTTTGGCGGCAGG
V  E  L  S  G  Q  K  H  V  F  M  F  G  A  M  G  F  G  G  R
      2230                2250                2270
GGCGCGTATGCCTTGGATTTAAGCAAAATCAACGGAAATTATCCGGCCGCCGCCCCCCTG
G  A  Y  A  L  D  L  S  K  I  N  G  N  Y  P  A  A  A  P  L
      2290                2310                2330
TTTGATGTCAAAGATGGCGATAATAACGGCAAAAATCGCGTGAAAGTGGAATTAGGCTAC
F  D  V  K  D  G  D  N  N  G  K  N  R  V  K  V  E  L  G  Y
      2350                2370                2390
ACCGTCGGTACGCCGCAAATCGGCAAAATCCGCAACGGCAAATACGCCGCCTTCCTCGCC
T  V  G  T  P  Q  I  G  K  I  R  N  G  K  Y  A  A  F  L  A
      2410                2430                2450
TCCGGTTATGCGGCTAAAAAAATTGACGACTCAACAAATAAAACCGCGCTGTATGTATAT
S  G  Y  A  A  K  K  I  D  D  S  T  N  K  T  A  L  Y  V  Y
      2470                2490                2510
GATTTGAAAGACACCTTAGGTACGCCGATTGCAAAAATCGAAGTGAAGGACGGCAAAGGC
D  L  K  D  T  L  G  T  P  I  A  K  I  E  V  K  D  G  K  G
```

FIG. 4-3a

```
      2530                2550                2570
GGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACGGTCGATATCGCCTAT
G  L  S  S  P  T  L  V  D  K  D  L  D  G  T  V  D  I  A  Y
      2590                2610                2630
GCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAATTCCGATTCTAGTAAA
A  G  D  R  G  G  N  M  Y  R  F  D  L  S  N  S  D  S  S  K
      2650                2670                2690
TGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCCGCGCCCGCCGTTTCC
W  S  A  K  V  I  F  E  G  D  K  P  I  T  S  A  P  A  V  S
      2710                2730                2750
CGACTGGCAGACAAACGCGTCGTCATCTTCGGTACGGGCAGCGATTTGACCGAAGATGAT
R  L  A  D  K  R  V  V  I  F  G  T  G  S  D  L  T  E  D  D
      2770                2790                2810
GTACTGAATACGGGCGAACAATATATTTACGGTATCTTTGACGACGATAAGGGGACGGTT
V  L  N  T  G  E  Q  Y  I  Y  G  I  F  D  D  D  K  G  T  V
      2830                2850                2870
AAGGTAACGGTACAAAACGGCACGGCAGGCGGGCTGCTCGAGCAACACCTTACTCAGGAA
K  V  T  V  Q  N  G  T  A  G  G  L  L  E  Q  H  L  T  Q  E
      2890                2910                2930
AATAAAACATTATTCCTGAACAAGAGATCCGACGGTTCGGGCAGCAAGGGCTGGGCGGTG
N  K  T  L  F  L  N  K  R  S  D  G  S  G  S  K  G  W  A  V
      2950                2970                2990
AAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACCGTGGTATTGCGTACCGCCTTC
K  L  R  E  G  E  R  V  T  V  K  P  T  V  V  L  R  T  A  F
      3010                3030                3050
GTAACCATCCGCAAATATAACGACGGCGGCTGCGGCGCGGAAACCGCCATTTTGGGCATC
V  T  I  R  K  Y  N  D  G  G  C  G  A  E  T  A  I  L  G  I
```

FIG. 4-3b

```
      3070                    3090                    3110
AATACCGCCGACGGCGGCGCATTGACTCCGAGAAGCGCGCGCCCGATTGTGCCGGATCAC
N  T  A  D  G  G  A  L  T  P  R  S  A  R  P  I  V  P  D  H
      3130                    3150                    3170
AATTCGGTTGCGCAATATTCCGGCCATAAGACAACCTCCAAAGGCAAATCCATCCCTATA
N  S  V  A  Q  Y  S  G  H  K  T  T  S  K  G  K  S  I  P  I
      3190                    3210                    3230
GGTTGTATGGACAAAGACGGTAAAACCGTCTGCCCGAACGGATATGTTTACGACAAGCCG
G  C  M  D  K  D  G  K  T  V  C  P  N  G  Y  V  Y  D  K  P
      3250                    3270                    3290
GTTAATGTGCGTTATCTGGATGAAACGGAAACAGACGGATTTTCAACGACGGCGGACGGC
V  N  V  R  Y  L  D  E  T  E  T  D  G  F  S  T  T  A  D  G
      3310                    3330                    3350
GATGCGGGCGGCAGCGGTATAGACCCCGCCGGCAGGCGTCCCGGCAAAAACAACCGCTGC
D  A  G  G  S  G  I  D  P  A  G  R  R  P  G  K  N  N  R  C
      3370                    3390                    3410
TTCTCCAAAAAAGGGGTGCGCACCCTGCTGATGAACGATTTGGACAGCTTGGATATTACC
F  S  K  K  G  V  R  T  L  L  M  N  D  L  D  S  L  D  I  T
      3430                    3450                    3470
GGCCCGATGTGCGGTATCAAACGCTTAAGCTGGCGCGAAGTCTTCTTCTGACCGGCCTGC
G  P  M  C  G  I  K  R  L  S  W  R  E  V  F  F  *
      3490                    3510                    3530
GCGGCCGGTTTTTCCGCAAATGCCGTCCGAAAGGCCTTCGGACGGCATTTTTTTGCGTTT
      3550
TTCGGGAGGGGGGCGGCAAATGAAACG
```

FIG. 5-1a

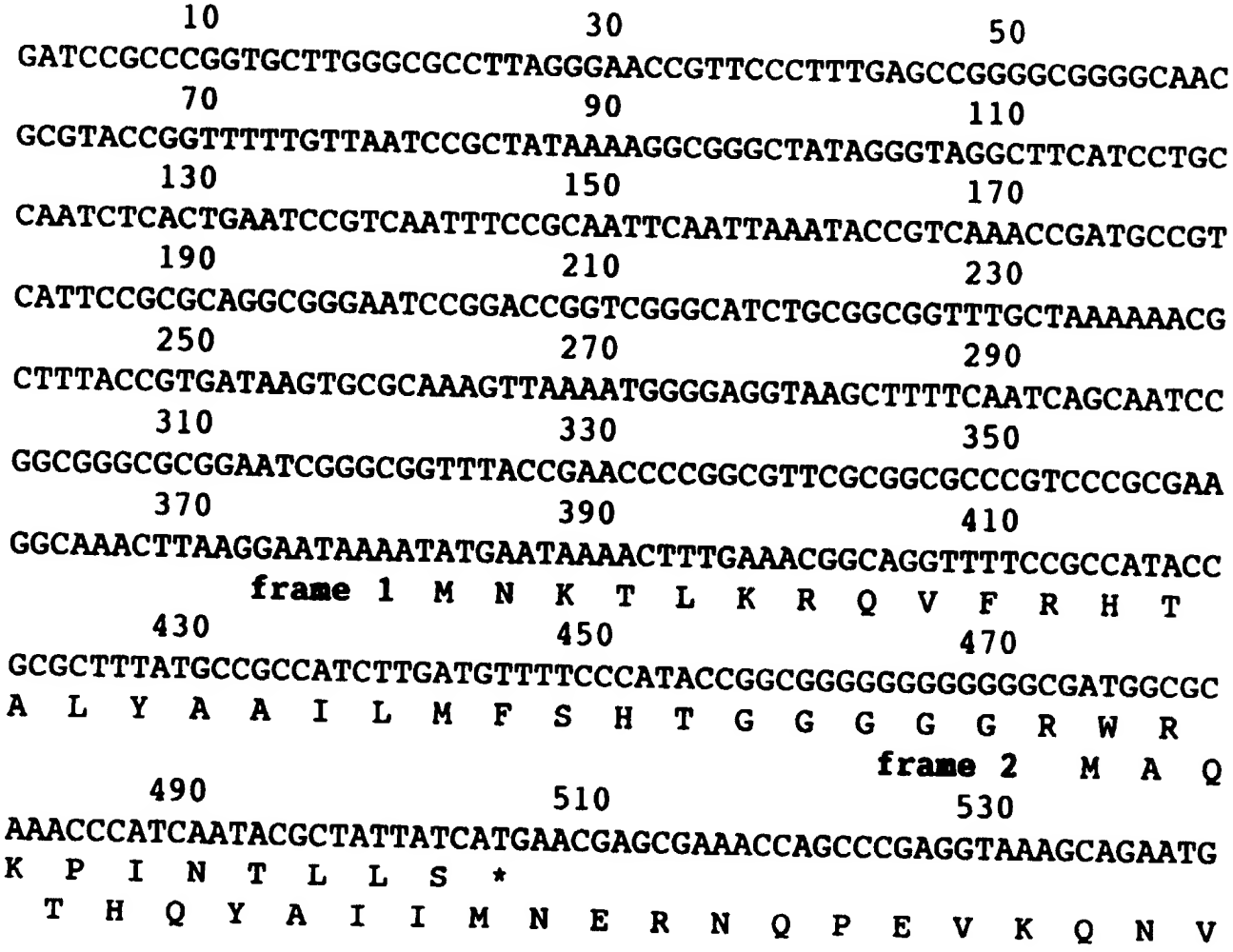

FIG. 5-1b

```
          550                 570                 590
TGCCATCTTCAATAAAGGACAAAGACAGGAGGCGCGAATATACTTATTATACGCACAGAA
  P  S  S  I  K  D  K  D  R  R  R  E  Y  T  Y  Y  T  H  R  T
          610                 630                 650
CAGGAGGAGGCTCTGTCTCATTCAACAATAACGATACCCTTGTTTCCCAACAAAGCGGTA
  G  G  G  S  V  S  F  N  N  N  D  T  L  V  S  Q  Q  S  G  T
          670                 690                 710
CTGCCGTTTTTGGCACAGCCACCTACCTGCCGCCCTACGGCAAGGTTTCCGGTTTTGATG
  A  V  F  G  T  A  T  Y  L  P  P  Y  G  K  V  S  G  F  D  A
          730                 750                 770
CCGTCGCTCTGAAAGAGCGCAACAATGCCGTTGATTGGATTCGTACCACCCGCATCGCGC
  V  A  L  K  E  R  N  N  A  V  D  W  I  R  T  T  R  I  A  L
          790                 810                 830
TGGCAGGCTACTCCTACATCGACGTCATATGCAGAAGCTACACAGGCTGTCCCAAACTTG
  A  G  Y  S  Y  I  D  V  I  C  R  S  Y  T  G  C  P  K  L  V
          850                 870                 890
TCTATAAAACCCGATTTACCTTCGGTCAACAAGGGTTGAAAAGAAAGGCAGGCAGCAAGC
  Y  K  T  R  F  T  F  G  Q  Q  G  L  K  R  K  A  G  S  K  L
          910                 930                 950
TGGATATATACGAAGACAAAAGCCGCGAAAATTCGCCCATTTACAAATTGTCGGATTATC
  D  I  Y  E  D  K  S  R  E  N  S  P  I  Y  K  L  S  D  Y  P
```

FIG. 5-1c

```
          970                       990                      1010
CTTGGTTGGGCGTATCTTTCAATTTGGGCAGCGAGAATACCGTCCAAAATAGCAAATTAT
  W  L  G  V  S  F  N  L  G  S  E  N  T  V  Q  N  S  K  L  F
         1030                      1050                     1070
TCAACAAATTGATATCTTCTTTTAGAGAAGGCAATAATAATCAAACCATCGTCTCTACGA
  N  K  L  I  S  S  F  R  E  G  N  N  N  Q  T  I  V  S  T  T
         1090                      1110                     1130
CAGAAGGCAACCCTATTTCCCTTGGCGACCGGCAGCGCGAACATACCGCCGTGGCCTATT
  E  G  N  P  I  S  L  G  D  R  Q  R  E  H  T  A  V  A  Y  Y
         1150                      1170                     1190
ATCTGAACGCCAAACTGCACCTGCTGGACAAAAAAGGGATTGAAGATATCGCCCAAGGCA
  L  N  A  K  L  H  L  L  D  K  K  G  I  E  D  I  A  Q  G  K
         1210                      1230                     1250
AAATAGTGGATTTGGGTATCTTGAAACCGCACGTCGAGACGACAGGACGAAGCTTGCTAG
  I  V  D  L  G  I  L  K  P  H  V  E  T  T  G  R  S  L  L  D
         1270                      1290                     1310
ATTTTTGGGCTAGGTGGGACATTAAAGATACCGGGCAGATTCCGGTCAAGCTCGGCCTGC
  F  W  A  R  W  D  I  K  D  T  G  Q  I  P  V  K  L  G  L  P
```

```
1330                1350                1370
CGCAAGTCAAAGCAGGCCGCTGCACCAACAAACCGAACCCCAATAATACCAAAGCCC
  Q  V  K  A  G  R  C  T  N  K  P  N  P  N  N  N  T  K  A  P
1390                1410                1430
CTTCGCCCGGCACTGACCGCCCCCGCGCTGTGGTTCGGACCCGGGCAAGATGGTAAGGCGG
  S  P  A  L  T  A  P  A  L  W  F  G  P  G  Q  D  G  K  A  E
1450                1470                1490
AGATGTATTCCGCTTCGGTTTCCACCTACCCCGACAGTTCGAGCAGCCGCATCTTCCTCC
  M  Y  S  A  S  V  S  T  Y  P  D  S  S  S  S  R  I  F  L  Q
1510                1530                1550
AAGAGCTGAAAACTCAAACCGAACCCGGCAAACCCGGCCGCTATTCCCTCAAATCTTTGA
  E  L  K  T  Q  T  E  P  G  K  P  G  R  Y  S  L  K  S  L  N
1570                1590                1610
ATGATGGTGAGATTAAAAGTCGACAGCCGAGTTTCAACGGGCGGCAAACAATCATCCGAT
  D  G  E  I  K  S  R  Q  P  S  F  N  G  R  Q  T  I  I  R  L
1630                1650                1670
TGGATGACGGCGTACATTTGATCAAACTGAATGGAAGCAAGGATGAGGTCGCCGCTTTTG
  D  D  G  V  H  L  I  K  L  N  G  S  K  D  E  V  A  A  F  V
1690                1710                1730
TCAATTTAAATGGAAACAACACCGGCAAAAACGACACTTTCGGCATTGTTAAGGAAGCGA
  N  L  N  G  N  N  T  G  K  N  D  T  F  G  I  V  K  E  A  N
1750                1770                1790
ACGTCAATCTTGACGCCGACGAGTGGAAAAAGTGCTGCCTTGGACGGTTCGGGGTC
  V  N  L  D  A  D  E  W  K  K  V  L  L  P  W  T  V  R  G  P
```

FIG. 5-2a

FIG. 5-2b

```
      1810                1830                1850
CCGATAATGACAATAAATTTAAATCAATTAACCAAAAACCAGAAAAATACAGCCAAAGAT
  D  N  D  N  K  F  K  S  I  N  Q  K  P  E  K  Y  S  Q  R  Y
      1870                1890                1910
ACCGCATCCGCGACAACAACGGCAATCGCGATTTGGGCGACATCGTCAACAGCCCGATTG
  R  I  R  D  N  N  G  N  R  D  L  G  D  I  V  N  S  P  I  V
      1930                1950                1970
TCGCGGTCGGCGGGTATTTGGCAACCGCCGCGAACGACGGGATGGTGCATATCTTCAAAA
  A  V  G  G  Y  L  A  T  A  A  N  D  G  M  V  H  I  F  K  K
      1990                2010                2030
AAAACGGCGGCAGTGATGAACGCAGCTACAATCTGAAGCTCAGCTACATCCCCGGCACGA
  N  G  G  S  D  E  R  S  Y  N  L  K  L  S  Y  I  P  G  T  M
      2050                2070                2090
TGCCGCGCAAGGATATTCAAAGCCAAGAATCCACCCTTGCCAAAGAGCTGCGCGCCTTTG
  P  R  K  D  I  Q  S  Q  E  S  T  L  A  K  E  L  R  A  F  A
      2110                2130                2150
CCGAAAAAGGCTATGTGGGCGACCGCTACGGCGTGGACGGCGGCTTTGTCTTGCGCCAAG
  E  K  G  Y  V  G  D  R  Y  G  V  D  G  G  F  V  L  R  Q  V
      2170                2190                2210
TCGAACTGAGCGGGCAAAAACACGTGTTTATGTTCGGCGCGATGGGTTTTGGCGGCAGGG
  E  L  S  G  Q  K  H  V  F  M  F  G  A  M  G  F  G  G  R  G
```

```
2230                2250                2270
GCGCGTATGCCTTGGATTTAAGCAAAATCAACGGAAATTATCCGGCCGCCGCCCCCCTGT
  A  Y  A  L  D  L  S  K  I  N  G  N  Y  P  A  A  A  P  L  F
     2290                2310                2330
TTGATGTCAAAGATGGCGATAATAACGGCAAAAATCGCGTGAAAGTGGAATTAGGCTACA
  D  V  K  D  G  D  N  N  G  K  N  R  V  K  V  E  L  G  Y  T
     2350                2370                2390
CCGTCGGTACGCCGCAAATCGGCAAAATCCGCAACGGCAAATACGCCGCCTTCCTCGCCT
  V  G  T  P  Q  I  G  K  I  R  N  G  K  Y  A  A  F  L  A  S
     2410                2430                2450
CCGGTTATGCGGCTAAAAAAATTGACGACTCAACAAATAAAACCGCGCTGTATGTATATG
  G  Y  A  K  K  I  D  D  S  T  N  K  T  A  L  Y  V  Y  D
     2470                2490                2510
ATTTGAAAGACACCTTAGGTACGCCGATTGCAAAAATCGAAGTGAAGGACGGCAAAGGCG
  L  K  D  T  L  G  T  P  I  A  K  I  E  V  K  D  G  K  G
     2530                2550                2570
GGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACGGTCGATATCGCCTATG
  L  S  S  P  T  L  V  D  K  D  L  D  G  T  V  D  I  A  Y  A
```

FIG. 5-2C

```
2590                2610                2630
CCGGCGACCGGGGGCGGCAATATGTACCGCTTTGATTTGAGCAATTCCGATTCTAGTAAAT
  G  D  R  G  G  N  M  Y  R  R  F  D  L  S  N  S  D  S  S  K  W
     2650                2670                2690
GGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCCGCGCCCGCCGTTTCCC
  S  A  K  V  I  F  E  G  D  K  P  I  T  S  A  P  A  V  S  R
     2710                2730                2750
GACTGGCAGACAAACGCGTCGTCATCTTCGGTACGGGCAGCGATTTGACCGAAGATGATG
  L  A  D  K  R  V  V  I  F  G  T  G  S  D  L  T  E  D  D  V
     2770                2790                2810
TACTGAATACGGGCGAACAATATATTTACGGTATCTTTGACGACGATAAGGGACGGTTA
  L  N  T  G  E  Q  Y  I  Y  G  I  F  D  D  D  K  G  T  V  K
     2830                2850                2870
AGGTAACGGTACAAAACGGCACGGCAGGCGGGCTGCTCGAGCAACACCTTACTCAGGAAA
  V  T  V  Q  N  G  T  A  G  G  L  L  E  Q  H  L  T  Q  E  N
     2890                2910                2930
ATAAAACATTATTCCTGAACAAGAGATCCGACGGTTCGGGCAGCAAGGGCTGGGCGGTGA
  K  T  L  F  L  N  K  R  S  D  G  S  G  S  K  G  W  A  V  K
     2950                2970                2990
AATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACCGTGGTATTGCGTACCGCCTTCG
  L  R  E  G  E  R  V  T  V  K  P  T  V  V  L  R  T  A  F  V
     3010                3030                3050
TAACCATCCGCAAATATAACGACGGCGGCTGCGGCGCGGAAACCGCCATTTTGGGCATCA
  T  I  R  K  Y  N  D  G  G  C  G  A  E  T  A  I  L  G  I  N
```

FIG. 5-3a

FIG. 5-3b

```
      3070                    3090                    3110
ATACCGCCGACGGCGGCGCATTGACTCCGAGAAGCGCGCGCCCGATTGTGCCGGATCACA
  T  A  D  G  G  A  L  T  P  R  S  A  R  P  I  V  P  D  H  N
      3130                    3150                    3170
ATTCGGTTGCGCAATATTCCGGCCATAAGACAACCTCCAAAGGCAAATCCATCCCTATAG
  S  V  A  Q  Y  S  G  H  K  T  T  S  K  G  K  S  I  P  I  G
      3190                    3210                    3230
GTTGTATGGACAAAGACGGTAAAACCGTCTGCCCGAACGGATATGTTTACGACAAGCCGG
  C  M  D  K  D  G  K  T  V  C  P  N  G  Y  V  Y  D  K  P  V
      3250                    3270                    3290
TTAATGTGCGTTATCTGGATGAAACGGAAACAGACGGATTTTCAACGACGGCGGACGGCG
  N  V  R  Y  L  D  E  T  E  T  D  G  F  S  T  T  A  D  G  D
      3310                    3330                    3350
ATGCGGGCGGCAGCGGTATAGACCCCGCCGGCAGGCGTCCCGGCAAAAACAACCGCTGCT
  A  G  G  S  G  I  D  P  A  G  R  R  P  G  K  N  N  R  C  F
      3370                    3390                    3410
TCTCCAAAAAAGGGGTGCGCACCCTGCTGATGAACGATTTGGACAGCTTGGATATTACCG
  S  K  K  G  V  R  T  L  L  M  N  D  L  D  S  L  D  I  T  G
      3430                    3450                    3470
GCCCGATGTGCGGTATCAAACGCTTAAGCTGGCGCGAAGTCTTCTTCTGACCGGCCTGCG
  P  M  C  G  I  K  R  L  S  W  R  E  V  F  F  *
      3490                    3510                    3530
CGGCCGGTTTTTCCGCAAATGCCGTCCGAAAGGCCTTCGGACGGCATTTTTTTGCGTTTT
      3550
TCGGGAGGGGGGCGGCAAATGAAACG
```

FIG. 6a

MS11 (PilC+)

136

1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGG

1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGG

2a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGG
His Thr Ala LeuTyr Ala Ala Ile Leu Met Phe SerHis ThrGly Gly Gly Gly GlyAla

2b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGG

UM01 (PilC+)

1c CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGG
His Thr Ala LeuTyr Ala Ala Ile Leu Met Phe SerHis ThrGly Gly Gly Gly GlyAla

1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGG

1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGG

765 (PilC+)

3a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGG
His Thr Ala LeuTyr Ala Ala Ile Leu Met Phe SerHis ThrGly Gly Gly Gly GlyAla

3b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGG

4 CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGCAG

605103 (PilC-)

1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGG

1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGG

pABJ04

1c CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGGG
His Thr Ala LeuTyr Ala Ala Ile Leu Met Phe SerHis ThrGly Gly Gly Gly GlyAla

1a CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGGG

1b CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGGG

1d CATACCGCGCTTTATGCCGCCATCTTGATGTTTTCCCATACCGGCGGGGGGGGGG
His Thr Ala LeuTyr Ala Ala Ile Leu Met Phe SerHis ThrGly Gly Gly GlyAla

FIG. 6b

	G-stretch	Sequenced clones
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	12 G	11
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	11 G	6
CGCAGGCGCAAACCCGTAAATACGCTATTATCATGAACGAGCGA Gln Ala Gln Thr Arg Lys Tyr Ala Ile Ile Met Asn Glu Arg 1 2 3 4 5 6 7 8 9 10 11 12	13 G	2
CGCAGGCGCAAACCCGTAAATACGCTATTATCATGAACGAGCGA	12 G	1
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA Met Ala Gln Thr His Gln Tyr Ala Ile Ile Met Asn Glu Arg	13 G	7
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	12 G	1
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	11 G	2
CGATGGCGCAAACGTATAAATACGCTATTGTGATGAACGAGCGA Met Ala Gln Thr Tyr Lys Tyr Ala Ile Val Met Asn Glu Arg	13 G	1
CGATGGCGCAAACGTATAAATACGCTATTGTGATGAACGAGCGA	14 G	5
GCGCAGGCGCAAACGTATAAATACGCTATTGTGATGAACGAGCGA	11 G	3
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	12 G	1
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	11 G	8
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA Met Ala Gln Thr His Gln Tyr Ala Ile Ile Met Asn Glu Arg	13 G	2
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	12 G	7
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA	11 G	1
CGATGGCGCAAACCCATCAATACGCTATTATCATGAACGAGCGA Met Ala Gln Thr His Gln Tyr Ala Ile Ile Met Asn Glu Arg	10 G	2

1 CCGCTGTATGTGTATGATTTGGAAAACACCAGTGGTAGTCTGATTAAAAAATCGAAGCA

CCCGGCGGCAAAGGCGGGCTTTCGTCCCCCCACGCTGGTGGATAAAGATTTGGACGGCACG

GTCGATATCGCCTATGCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAAT

TCCGATTCTAGTAAATGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCC

GCGCCCGCCGTTTCCCGACTGGCAGACAAACGCGTGGTTATCTTCGGCACGGGCAGCGAT

TTGAGTGAACAGGATGTACTGGATACGGACAAACAATATATTTACGGTATCTTTGACGAC

GATAAGTCGACGGTTAATGTAAAGGTAACAAACGGCACGGGAGGCGGGCTGCTCGAGCAA

GTGCTTAAAGAGGAAAGTAAAACCTTATTCCTGAGCAATAATAAGGCATCCGGCGGATCG

GCCGATAAAGGGTGGGTAGTGAAATTGAGGGAAGAGAACGCGTTACCGTCAAACCGACC

GTGGTATTGCGTACCGCCTTTGTCACCATCCGCAAATATACGGATACGACAAATGTGGC

GCGCAAACCGCCATTTTGGGCATCAATACCGCCCGACGGCGGCGCATTGACTCCGAGAAGC

GCGCGCCCGATTGTGCCGGATCACAATTCGGTTGCGCAATATTCCGGCCATCAGAAAATG

AACGGCAAGTCCATCCCGG 739

FIG. 7

```
   P  L  Y  V  Y  D  L  E  N  T  S  G  S  L  I  K  K  I  E  A
1  CCGCTGTATGTGTATGATTTGGAAAACACCAGTGGTAGTCTGATTAAAAAAATCGAAGCA

   P  G  G  K  G  G  L  S  S  P  T  L  V  D  K  D  L  D  G  T
   CCCGGCGGCAAAGGCGGGCTTTCGTCCCCCACGCTGGTGGATAAAGATTTGGACGGCACG

   V  D  I  A  Y  A  G  D  R  G  G  N  M  Y  R  F  D  L  S  N
   GTCGATATCGCCTATGCCGGCGACCGGGGCGGCAATATGTACCGCTTTGATTTGAGCAAT

   S  D  S  S  K  W  S  A  K  V  I  F  E  G  D  K  P  I  T  S
   TCCGATTCTAGTAAATGGTCTGCAAAGGTTATTTTCGAAGGCGACAAGCCGATTACCTCC

   A  P  A  V  S  R  L  A  D  K  R  V  V  I  F  G  T  G  S  D
   GCGCCCGCCGTTTCCCGACTGGCAGACAAACGCGTGCTTATCTTCGGCACGGGCAGCGAT

   L  S  E  Q  D  V  L  D  T  D  K  Q  Y  I  Y  G  I  F  D  D
   TTGAGTGAACAGGATGTACTGGATACGGACAAACAATATATTTACGGTATCTTTGACGAC

   D  K  S  T  V  N  V  K  V  T  N  G  T  G  G  G  L  L  E  Q
   GATAAGTCGACGGTTAATGTAAAGGTAACAAACGGCACGGGAGGCGGGCTGCTCGAGCAA
```

FIG. 8-1

SUBSTITUTE SHEET

```
V L K E E S K T L F L S N N K A S G G S
GTGCTTAAAGAGGAAAGTAAAACCTTATTCCTGAGCAATAAGGCATCCGGCGGATCG

A D K G W V V K L R E G E R V T V K P T
GCCGATAAAGGGTGGTAGTGAAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACC

V V L R T A F V T I R K Y T D T D K C G
GTGGTATTGCGTACCGCCTTTGTCACCATCCGCAAATATACGGATACGGACAAATGTGGC

A Q T A I L G I N T A D G G A L T P R S
GCGCAAACCGCCATTTTGGGCATCAATACCGCCGACGGCGGCGCATTGACTCCGAGAAGC

A R P I V P D H N S V A Q Y S G H Q K M
GCGCGCCCGATTGTGCCGGATCACAATTCGGTTGCGCAATATTCCGGCCATCAGAAAATG.

N G K S I P
AACGGCAAGTCCATCCCGG  739
```

FIG. 8-2

FIG. 9-1a

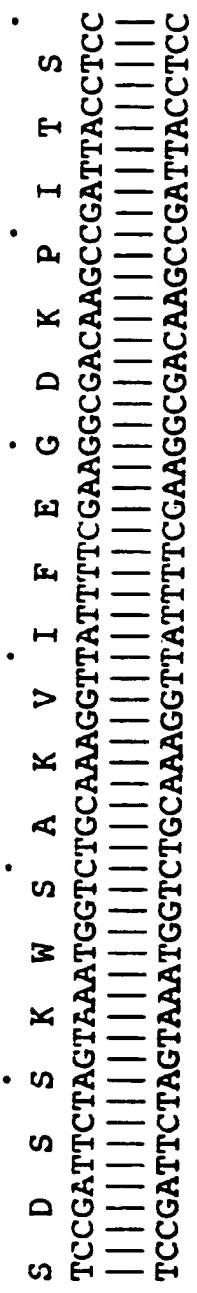

FIG. 9-1b

FIG. 9-1c

```
D  K  S  T  V  N  V  K  V  T  N  G  T  G  G  G  L  L  E  Q
GATAAGTCGACGGTTAATGTAAAGGTAACAAACGGCACGGGAGGCGGGCTGCTCGAGCAA
||||||  |||||||||  ||||  ||||   |||||||||||  |||||||||||||||||||
GATAAGGGGACGGTTAAGGTAACGGTACAAAACGGCACGGCAGGCGGGCTGCTCGAGCAA
D  K  G  T  V  K  V  T  V  Q  N  G  T  A  G  G  L  L  E  Q

V  L  K  E  E  S  K  T  L  F  L  S  N  N  K  A  S  G  G  S
GTGCTTAAAGAGGAAAGTAAAACCTTATTCCTGAGCAATAATAAGGCATCCGGCGGATCG
   ||||   |||||| |||||| |||||||||      || |||  ||||| ||| |||
CACCTTACTCAGGAAAATAAAACATTATTCCTG......AACAAGAGATCCGACGGTTCG
H  L  T  Q  E  N  K  T  L  F  L        N  K  R  S  D  G  S

A  D  K  G  W  V  V  K  L  R  E  G  E  R  V  T  V  K  P  T
GCCGATAAAGGGTGGGTAGTGAAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACC
| |   || || ||||  ||||||||||||||||||||||||||||||||||||||||
GGCAGCAAGGGCTGGGCGGTGAAATTGAGGGAAGGAGAACGCGTTACCGTCAAACCGACC
G  S  K  G  W  A  V  K  L  R  E  G  E  R  V  T  V  K  P  T
```

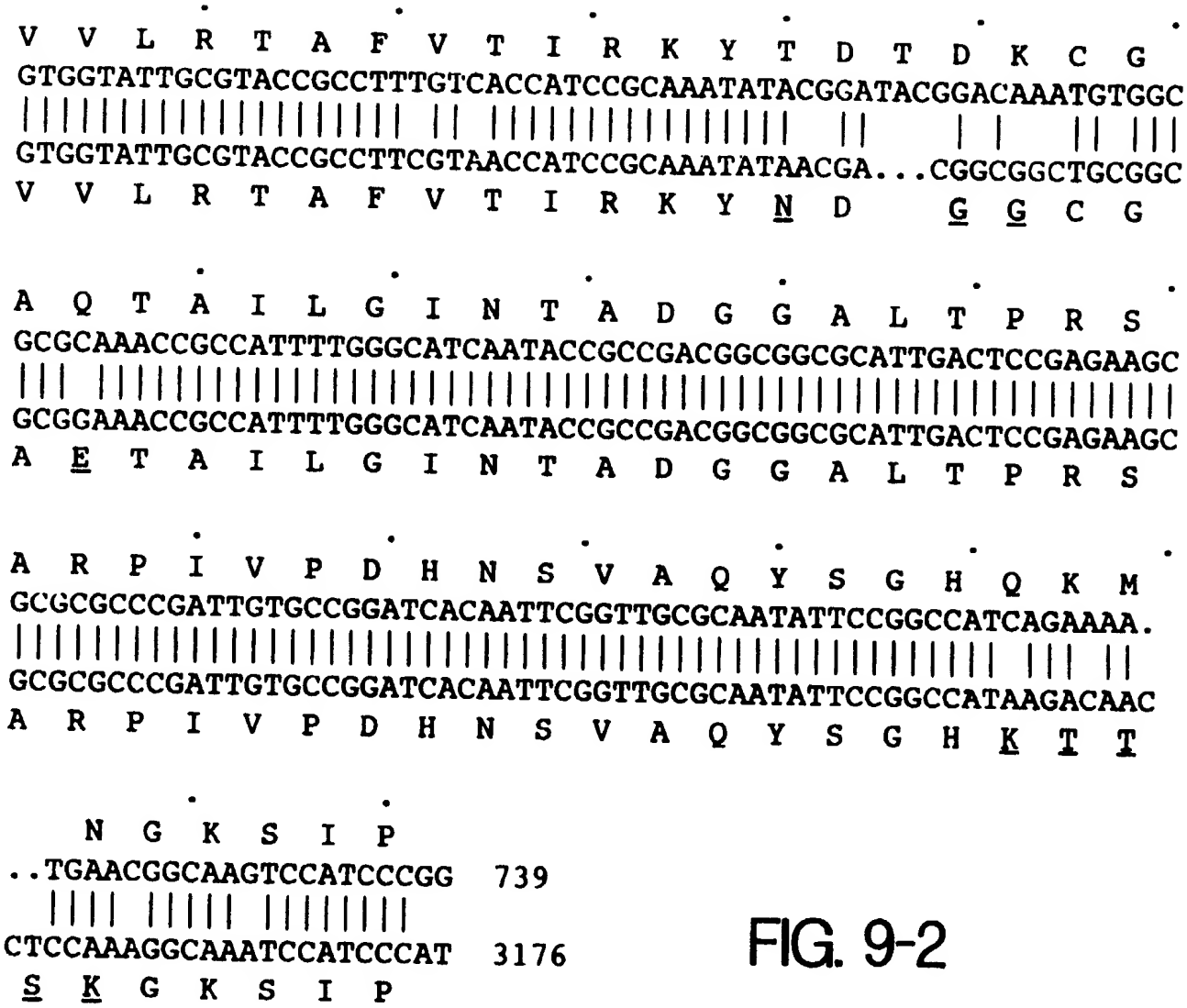

FIG. 9-2

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US92/00863

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C07H 21/00; G01N 33/53
US CL : 536/27; 435/7.3

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	536/27-29; 435/6,7.1,7.2,7.3,7.36,69.1, 69.3,91,243,871,961; 436/501,519,547,811; 530/300,324,333,350,388 4,389.5,403,825: 935/9, 12, 15, 66, 72, 73, 91

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

CAS ONLINE, MEDLINE, APS, BIOSIS

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
X/Y	US,A, 4,443,431 (BUCHANAN ET AL.) 17 APRIL 1984, SEE ESPECIALLY COLUMN 3, LINES 1-7.	1 - 4 , 7 - 10,33,34,38/5, 6 , 1 1 , 1 2 - 24,26,30,37,41
A	JOURNAL OF BACTERIOLOGY, VOLUME 130, NUMBER 1, ISSUED APRIL 1977, SWANEY ET AL., "GENETIC COMPLEMENTATION ANALYSIS OF ESCHERICHIA COLI TYPE 1 SOMATIC PILUS MUTANTS", PAGES 506-511, SEE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41
X/Y	US,A, 4,584,195 (SCHOOLNIK ET AL.) 22 APRIL 1986, SEE ESPECIALLY THE ABSTRACT AND CLAIMS 1-14.	1 - 4 , 7 - 10,22,23/5,6,1 1,12,24,30,33, 34,37,38,41
Y	JOURNAL OF BACTERIOLOGY, VOLUME 172, NUMBER 6, ISSUED JUNE 1990, NUNN ET AL., "PRODUCTS OF THREE ACCESSORY GENES, PILB, PILC, PILD, ARE REQUIRED FOR BIOGENESIS OF PSEUDOMONAS AERUGINOSA PILI", PAGES 2911-2919, SEE ESPECIALLY THE ABSTRACT.	1 - 24,26,30,33,34 ,37,38, 41

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
27 MAY 1992	02 JUN 1992
International Searching Authority[1]	**Signature of Authorized Officer[20]**
ISA/US	ARDIN H. MARSCHEL

Form PCT/ISA/210 (second sheet)(May 1986) ©

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	JOURNAL OF BACTERIOLOGY, VOLUME 170, NUMBER 4, ISSUED APRIL 1988, PERRY ET AL., "NEISSERIA MENINGITIDIS C114 CONTAINS SILENT, TRUNCATED PILIN GENES THAT ARE HOMOLOGOUS TO NEISSERIA GONORRHOEAE PIL SEQUENCES", PAGES 1691-1697, SEE ESPECIALLY PAGE 1691, SECOND COLUMN, LINES 3-8.	1 24,26,30,33,34 ,37,38,41
X/Y	THE JOURNAL OF BIOLOGICAL CHEMISTRY, VOLUME 254, NUMBER 9, ISSUED 10 MAY 1979, KELLEY ET AL., "A RAPID PROCEDURE FOR ISOLATION OF LARGE QUANTITIES OF ESCHERICHIA COLI DNA POLYMERASE I UTILIZING A LAMBDA-POL A TRANSDUCING PHAGE", PAGES 3206-3210, SEE ESPECIALLY TABLE II ON PAGE 3208.	12,24/30,41

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE[1]

1. ☐ Claim numbers _, because they relate to subject matter (1) not required to be searched by this Authority, namely:

2. ☐ Claim numbers _, because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out (1), specifically:

3. ☐ Claim numbers _, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING[2]

This International Searching Authority found multiple inventions in this international application as follows:

Please See Attached Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

1-24,26,30,33,34,37,38,41 (Telephone Practice) (Telephone Practice) (Telephone Practice)

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Search Authority did not invite payment of any additional fee.

Remark on protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No.[18]
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (USA), VOLUME 84, ISSUED DECEMBER 1987, HAAS ET AL., "RELEASE OF SOLUBLE PILIN ANTIGEN COUPLED WITH GENE CONVERSION IN NEISSERIA GONORRHOEAE", PAGES 9079-9083, SEE THE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41
X/Y	NEW ENGLAND BIOLABS CATALOG, ISSUED 1986, (NEW ENGLAND BIOLABS, BEVERLY, MASSACHUSETTS, 1986), PAGE 60, SEE ESPECIALLY LINKER # 1096 COMPARED TO THE INSTANT APPLICATION FIGURE 3 AT BASES 3196-3202.	33,34/37
X/Y	SIGMA CHEMICAL COMPANY CATALOG, ISSUED 1990, (SIGMA CHEMICAL COMPANY, ST. LOUIS, MISSOURI, 1990), PAGES 859-860, SEE ESPECIALLY POLY[C]-[dG]12-18 ON PAGE 859 AND POLYDEOXYGUANYLIC ACID ON PAGE 860 COMPARED TO THE INSTANT APPLICATION IN FIGURE 3 AT BASES 461-472.	33,34,38/37
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES (USA), VOLUME 81, ISSUED OCTOBER 1984, MEYER ET AL., "PILUS GENES OF NEISSERIA GONORRHEAE: CHROMOSOMAL ORGANIZATION AND DNA SEQUENCE", PAGES 6110-6114, SEE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41
A	JOURNAL OF GENERAL MICROBIOLOGY, VOLUME 132, ISSUED 1986, TINSLEY ET AL., "VARIATION IN THE EXPRESSION OF PILI AND OUTER MEMBRANE PROTEIN BY NEISSERIA MENINGITIDIS DURING THE COURSE OF MENINGOCOCCAL INFECTION", PAGES 2483-2490, SEE ENTIRE DOCUMENT.	1 - 24,26,30,33,34 ,37,38,41